10617443

10/617,443

Apologies!!

-----Original Message-----
From: STIC-Biotech/ChemLib
Sent: Thursday, June 23, 2005 6:33 AM
To: Swope, Sheridan
Subject: RE: 10/617,433

Good morning, do you have the correct number. 10/617433 is NO BIOTECH DATA.

-----Original Message-----
From: Swope, Sheridan
Sent: Wednesday, June 22, 2005 5:25 PM
To: STIC-Biotech/ChemLib
Subject: 10/617,433

For 10/617,433, pls search:

SID 1: residues 1-1038 oligo search (≥12 nts) against the NT databases.

SID 1: residues 1-1038 regular search against the NT databases.

SID 2: residues 1-9 against the NT databases.

SID 2: regular search against the NT databases.

Thanks!!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
571-272-0943 (voice)
E02B71 Remsen Bld (Office)
E02C70 Remsen Bld (Mailbox)

Barb O'Bryen

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: June 29, 2005, 13:51:28 ; Search time 4887 Seconds
(without alignments)
10291.900 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 cagggactcgaagttgcag.........ggcccttcccgccagcgcag 1038

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum Match 0%
Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_btg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1031.6	99.4	60597	9	AC113611	Homo sapi
2	83	8.0	218856	2	AC122871	Mus muscu
3	77.2	7.4	235102	2	AC118993	Rattus no
4	77.2	7.4	258368	2	AC110369	Rattus no
5	77.2	7.4	264075	2	AC134075	Rattus no
6	56.4	5.4	125020	2	AF429315	Homo sapi
7	56.2	5.4	172105	2	AC021197	Homo sapi
8	55.8	5.4	98501	9	AP005019	Homo sapi
9	55.8	5.4	111282	6	AP005019	Homo sapi
10	55.8	5.4	111282	6	AR368084	Sequence
11	55.8	5.4	194719	2	AX662038	Sequence
12	55.8	5.4	194719	2	AC026173	Homo sapi
13	55.8	5.4	197022	2	AC055829	Homo sapi
14	54.8	5.3	125020	9	AF429315	Homo sapi
15	53.4	5.1	173791	9	AC021035	Homo sapi
16	53.2	5.1	196512	2	AC020911	Homo sapi
17	52.8	5.1	192138	2	AC148545	Otolemur
18	52.4	5.0	147454	9	AC027348	Homo sapi
19	52.4	5.0	126380	2	AP000714	Homo sapi
		5.0	159840	2	AP000869	Homo sapi
20	52.4	5.0	171902	2	AP000846	Homo sapi
21	52.4	5.0	197215	2	AP002954	Homo sapi
22	52.4	5.0	199321	2	AP000941	Homo sapi
23	52.2	5.0	90801	9	AC025033	Homo sapi
24	52.2	5.0	171615	2	AC089988	Homo sapi
25	52.2	5.0	212382	2	AC010936	Homo sapi
26	52	5.0	102216	9	AC010936	Homo sapi
27	51	4.9	14725	9	AL161908	Human DNA
28	51	4.9	112560	9	Z68881	Human DNA s
29	51	4.9	125111	2	AL109760	Human DNA
30	51	4.9	212406	2	AC147938	Ornithorh
31	51	4.9	217418	2	AC132705	Homo sapi
32	51	4.9	244546	2	AC106071	Rattus no
33	50.8	4.9	102591	9	AC107358	Rattus no
34	50.8	4.9	53269	2	AC002978	Homo sapi.
35	50.6	4.9	75031	9	AC006184	Homo sapi.
36	50	4.9	63266	9	AL390716	Human DNA
37	50	4.8	183672	9	AL607038	Human chr
38	50	4.8	122251	9	AL355074	Human chr
39	49.6	4.8	118495	9	AL161662	Human DNA
40	49.6	4.8	122251	9	AL590556	Human DNA
41	49.6	4.8	162805	2	AP003780	Homo sapi
42	49.6	4.8	192453	2	AC020565	Homo sapi
43	49.6	4.8	2000	2	AX655393	Sequence
44	49.4	4.8	183309	6	AC150795	Bos tauru
45	49.2	4.7	112911	9	AL355310	Human DNA

ALIGNMENTS

RESULT 1
AC113611
AC113611
LOCUS   AC113611   60597 bp   DNA   linear   PRI 10-MAR-2003
DEFINITION   Homo sapiens BAC clone RP11-421M20 from 4, complete sequence.
ACCESSION   AC113611
VERSION   AC113611.3 GI:28475761
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 60597)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE   Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED   9847074
REFERENCE   2 (bases 1 to 60597)
AUTHORS   Isak,A. and Cotton,M.
TITLE   The sequence of Homo sapiens BAC clone RP11-421M20
JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 60597)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (04-MAR-2002) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   4 (bases 1 to 60597)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (11-APR-2002) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   5 (bases 1 to 60597)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   6 (bases 1 to 60597)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission

JOURNAL   Submitted (10-MAR-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Feb 23, 2003 this sequence version replaced gi:20136967.
COMMENT   -------------- Genome Center
                   Center: Washington University Genome Sequencing Center
                   Center code: WUGSC
                   Web site: http://genome.wustl.edu/gsc
                   Contact: sapiens@watson.wustl.edu

                   Summary Statistics
                   Center project name: H_NH0421M20
           --------------.

           NOTICE: This sequence may not represent the entire insert of this
           clone. It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded, sequenced with an alternate
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

           MAPPING INFORMATION:
           Mapping information for this clone was provided by Dr. John D.
           McPherson, Department of Genetics, Washington University, St. Louis
           MO. For additional information about the map position of this
           sequence, see http://genome.wustl.edu/gsc

           SOURCE INFORMATION:
           The RPCI-11 human BAC library was made from the blood of one male
           donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
           Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
           approach for construction of bacterial artificial chromosome
           libraries. Genomics 51:1-8. The clone may be obtained either from
           Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
           and coworkers at http://www.chori.org
           VECTOR: pBACe3.6

           NEIGHBORING SEQUENCE INFORMATION:
           This sequence is not the entire insert of the clone. This clone is
           overlapped by AC104650 and AC104825.

           Discrepant bases between AC104650, AC104825 and clone sequence.

           Data from AC104825 was used to finish this clone.
FEATURES   Location/Qualifiers
     source   1..60597
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-421M20"
                     /clone_lib="RPCI-11"
     repeat_region   539..903
                     /rpt_family="ERVL"
     repeat_region   2797..2907
                     /rpt_family="(CAGAGA)n"
     repeat_region   2908..3114
                     /rpt_family="ERVL"
     repeat_region   3122..3184
                     /rpt_family="(CAGAGA)n"
     repeat_region   3256..3368
                     /rpt_family="L1"
     repeat_region   3500..3682
                     /rpt_family="L1"
     repeat_region   3713..3782
                     /rpt_family="AT_rich"
     repeat_region   5208..5410
                     /rpt_family="MIR"

     repeat_region   5451..5503
                     /rpt_family="MIR"
     repeat_region   5504..5677
                     /rpt_family="MIR"
     repeat_region   6645..6687
                     /rpt_family="MIR"
     repeat_region   6998..7067
                     /rpt_family="MIR"
     repeat_region   7741..8173
                     /rpt_family="MIR"
     repeat_region   8530..8610
                     /rpt_family="MER1_type"
     misc_feature   8638..9514
                     /note="CpG_island (%GC=76.7, o/e=0.80, #CpGs=114)"
     repeat_region   9027..9048
                     /rpt_family="GC_rich"
     repeat_region   9063..9111
                     /rpt_family="GC_rich"
     repeat_region   9612..9739
                     /rpt_family="GC_rich"
     repeat_region   10414..10509
                     /rpt_family="MIR"
     repeat_region   10688..11034
                     /rpt_family="MER103"
     repeat_region   11781..11997
                     /rpt_family="MaLR"
     repeat_region   12212..12325
                     /rpt_family="L2"
     repeat_region   13275..13625
                     /rpt_family="MIR"
     repeat_region   13703..13769
                     /rpt_family="L2"
     repeat_region   14926..15054
                     /rpt_family="CR1"
     repeat_region   15239..15281
                     /rpt_family="MIR"
     repeat_region   15349..15515
                     /rpt_family="Tip100"
     repeat_region   16000..16113
                     /rpt_family="MIR"
     repeat_region   16639..16917
                     /rpt_family="MIR"
     repeat_region   17127..17228
                     /rpt_family="L2"
     repeat_region   17883..17982
                     /rpt_family="Tip100"
     repeat_region   18108..18173
                     /rpt_family="MIR"
     repeat_region   19059..19127
                     /rpt_family="CT-rich"
     repeat_region   20087..20234
                     /rpt_family="L2"
     repeat_region   20576..20738
                     /rpt_family="MIR"
     repeat_region   22242..22280
                     /rpt_family="MIR"
     repeat_region   23733..23804
                     /rpt_family="(TGAA)n"
     repeat_region   23890..23987
                     /rpt_family="L2"
     repeat_region   24011..24440
                     /rpt_family="MIR"
     repeat_region   25254..25412
                     /rpt_family="MaLR"
     repeat_region   26001..26150
                     /rpt_family="MIR"
     repeat_region   26473..26512
                     /rpt_family="MIR"
     repeat_region   26513..26894
                     /rpt_family="MIR"
     repeat_region   26895..27048
                     /rpt_family="MaLR"
     repeat_region   27423..27724
                     /rpt_family="MIR"

```
Qy    repeat_region    27725 . 28016    /rpt_family="Alu"
Db    repeat_region    28029 . 28096    /rpt_family="Alu"
Db    repeat_region    28099 . 28399    /rpt_family="(TTTC)n"
Db    repeat_region    28406 . 28597    /rpt_family="Alu"
Db    repeat_region    29613 . 29783    /rpt_family="MER1_type"
Db    repeat_region    29988 . 30078    /rpt_family="L1"
Qy    repeat_region    30801 . 30893    /rpt_family="MIR"
Db    repeat_region

Query Match    99.4%;    Score 1031.6;    DB 9;    Length 60597;
Best Local Similarity    99.6%;    Pred. No. 1.3e-201;
Matches 1034;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;
```

```
Qy        1 CAGGGACTCGAAGTTTGCAGTTGCAGTTCCTCCACACTCAGTTCCCACAGATGTGGTAGGAGGGCAT       60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    20549 CAGGGACTCGAAGTTTGCAGTTGCAGTTCCTCCACACTCAGTTCCCACAGATGTGGTAGGAGGGCAT    20608

Qy       61 ATTCAGTCCCATTTTTCAGATGAGGAAACGTAAGTAATTCGTCTGA      120
            ||||||||||||||||||||||||||||||||||||||||||||||
Db    20609 ATTCAGTCCCATTTTTCAGATGAGGAAACGTAAGTAATTCGTCTGA     20668

Qy      121 GGCCACACAGCTAGAAAAGCAGCAGCCCAGCCCAGCCCAGC      180
            |||||||||||||||||||||||||||||||||||||||||
Db    20669 GGCCACACAGCTAGAAAAGCAGCAGCCCAGCCCAGCCCAGC     20728

Qy      181 CCAGTTGCTCATTGCGGGGGCTCGGGGGAGCCCACGGGGGC      240
            |||||||||||||||||||||||||||||||||||||||||
Db    20729 CCAGTTGCTCATTGCGGGGGCTCGGGGGAGCCCACGGGGGC     20788

Qy      241 TGGTGGGAACTGGAGAGGAGGCCCCGTGTCAGGGAACCCCG      300
            |||||||||||||||||||||||||||||||||||||||||
Db    20789 TGGTGGGAACTGGAGAGGAGGCCCCGTGTCAGGGAACCCCG     20848

Qy      301 CCCGTGCCACTGCAATGCCTCAGGCCTGTGGTCCTCTGCAG      360
            |||||||||||||||||||||||||||||||||||||||||
Db    20849 CCCGTGCCACTGCAATGCCTCAGGCCTGTGGTCCTCTGCAG     20908

Qy      361 GCAGGCAGGAAATCTGAGCCCCGGGAAGGGTCCAGGGAGTTCG      420
            |||||||||||||||||||||||||||||||||||||||||||
Db    20909 GCAGGCAGGAAATCTGAGCCCCGGGAAGGGTCCAGGGAGTTCG    20968

Qy      421 GGCTCGGGGTGTGGCCAAGTTAGACACAGATGTAGGGCCCCTG      480
            |||||||||||||||||||||||||||||||||||||||||||
Db    20969 GGCTCGGGGTGTGGCCAAGTTAGACACAGATGTAGGGCCCCTG    21028

Qy      481 CTCTTTTTGGCCCAGAGGGGCCACACTGTGTCCGGGGGCCGCA      540
            |||||||||||||||||||||||||||||||||||||||||||
Db    21029 CTCTTTTTGGCCCAGAGGGGCCACACTGTGTCCGGGGGCCGCA    21088

Qy      541 GGGGGTCTTCCACTACACCCCCG      600
            |||||||||||||||||||||||
Db    21089 GGGGGTCTTCCACTACACCCCCG     21148

Qy      601 AGCTGGAGCCCATGAGGAGGAGG      660
            |||||||||||||||||||||||
Db    21149 AGCTGGAGCCCATGAGGAGGAGG     21208

Qy      661 GGAACAGACAAAGGCCCAGGGGA      720
            |||||||||||||||||||||||
Db    21209 GGAACAGACAAAGGCCCAGGGGA     21268

Qy      721 GCTCACGGCAAATATCCTAACCT      780
            |||||||||||||||||||||||
Db    21269 GCTCACGGCAAATATCCTAACCT     21328

Qy      781 AGGGGGGTGAGGAACCCCACGTG      840
            |||||||||||||||||||||||
Db    21329 AGGGGGGTGAGGAACCCCACGTG     21388

Qy      901 AGTGTGGGCTGAGGGCGTGGCGT      960
            |||||||||||||||||||||||
Db    21449 AGTGTGGGCTGAGGGCTGGCGT     21508

Qy      961 TGCCCACTGGGCCATAGGCGCCT     1020
            |||||||||||||||||||||||
Db    21509 TGCCCACTGGGCCATAGGCGCCT    21568

Qy     1021 CCCTTCCCGCGGCCAGGCGCAG     1038
            |||||||||||||||||||||
Db    21569 CCCTTCCCGCGGCCAGGCGCAG    21586
```

```
RESULT 2
LOCUS       AC122871           218856 bp    DNA     linear   HTG 08-AUG-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-107P12, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC122871
VERSION     AC122871.3  GI:33504638
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 218856)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218856)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 218856)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            On Aug 8, 2003 this sequence version replaced gi:28461106.
COMMENT     -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information --------------
            Center project name: M_BA0107P12
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 217551 bases at least Q40
            Consensus quality: 217980 bases at least Q30
            Consensus quality: 218333 bases at least Q20
            Insert size: 198000; agarose-fp
            Insert size: 220194; sum-of-contigs
            Quality coverage: 15.43 in Q20 bases; agarose-fp
            Quality coverage: 13.45 in Q20 bases; sum-of-contigs
            --------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.

FEATURES             Location/Qualifiers
     source          1..218856
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-107P12"
     misc_feature    1..1554
                     /note="assembly_name:Contig9"
     misc_feature    1655..2784
                     /note="assembly_name:Contig15"
     misc_feature    2885..4046
                     /note="assembly_name:Contig17"
     misc_feature    4147..121446
                     /note="assembly_name:Contig20"
     misc_feature    121547..218856
                     /note="assembly_name:Contig21"

ORIGIN
        1  1554  contig of 1554 bp in length
     1555  1654:  gap of unknown length
     1655  2784:  contig of 1130 bp in length
     2785  2884:  gap of unknown length
     2885  4046:  contig of 1162 bp in length
     4047  4146:  gap of unknown length
     4147  121446:  contig of 117300 bp in length
   121447  121546:  gap of unknown length
   121547  218856:  contig of 97310 bp in length

Matches 295;  Conservative   0;  Mismatches 160;  Indels  36;  Gaps   8;

Query Match       8.0%;  Score 83;  DB 2;  Length 218856;
Best Local Similarity  60.1%;  Pred. No. 1.7e-07;

Qy    53 GAGGGCATATTCAGTCCCATTTTTCACGATGAG-GAGTTGAGCGCCCAGAGAACGTAAGTAA 111
         ||||| |  |||||| |||||   |||| | ||  ||||||| | |||||| ||| ||| ||
Db 200999 GGGGTCACCTTCAGTCCCATTGTTTAGATGGGAAAGTCAAGCGCCCAGAGAGGGGTAACTGA 201058

Qy   112 TCTGTCTGAGGCCACACAGCTAGAAAAGCAGCCGAAACCCCTGGTGTGTGTGTGCA 171
         ||||     ||| || ||||| ||| | ||  ||||| ||||||||| || |||||
Db 201059 CTTGTTCACAGGTCACACAGCTAAAAGACAGAGCAGGTCCAGTCAGG--CCTAGTTGTGTGTGCA 201116

Qy   172 GCCCCAGCCCAGTTGCTCATTCCGGGGGCTCGGAGCCCACGACGGAGGCTGACGAGCATG 231
         ||| ||||||||| |||||||||| |||||||||| | ||| |||| |||| || | |||
Db 201117 GCTCCCAGGCCTG-TGCTCATTTAGACTCCAGGAACCTCCATAAGGCACTGCCAGGCAG 201175

Qy   232 TGTTCCAGATGGTGGGAACTGAGAGAGCCCCGCGCACAGGGCCCGTGCAGGGAACCCCGAGG 291
         ||||| ||| || ||||        |||||||||||
Db 201176 TGTG------------------GTCACCGAGACCGGGACCTGGAGGTCTC------ 201207

Qy   292 GCTGCAGTGGCAGGGCCCCGTGCCACTGCATGCCTCAGGCGCTGCTCCAGCCCACAGCGCCCTA 351
         |||| ||||   || ||||||| |||| ||| | |||| ||  ||||| ||| ||| |||
Db 201208 AGTTTAGGCTCATGTCATCCACACATGCCTCTCGGGCCCTGTGGGTCTCTCGGCAAGCCACAGCCCCTG 201267

Qy   352 CTGCTGACGGCACCAGGAATCTCAGCCCCGGGAAG-GGTCCAGGAGAAGTTCGTGAACCATC 410
         || | | |  | |||| ||| |||||| ||||||  ||||||||| || |||||  ||| |
Db 201268 TTGCTGATGGCGGCAGAAATCTGAGCCACTGATGAGCGCTTTGGGGCCGTTCGTGTGCTGCC 201327

Qy   411 TAGCAGCTCGGCGCCTCGGGGTGTGGCCAGTTAGACACAGAATGAGTAGGGCCCTGTGGGACTCAG 470
         |||||| |||||  | |||||| || ||| |  | ||| ||| || || ||| ||| |||| || |
Db 201328 TGGCAAGTAGGCCCTTGGGCCAACGAGTTGGACAAAG--GCCGGTGTTCTGTCCACTCAG 201385

Qy   471 AAATTGGCAGCTCTTTTGCCCAGAGGGGGCCACGGCTGTGTCCGGGGGGTTAGCTCAGA 530
         |||  |   |||| |  |   ||| ||||||  || || |  |||| |||| |||| |||
Db 201386 GAATCTATAACCCACGGCCAAATGAGGAAGGGGCCTTCACGCAGGCC-AGCTGAGATGGCTCAGG 201444

Qy   531 AGGGTCACCTG 541
         |||| |||| |
Db 201445 AGGGGCCCCAG 201455
```

```
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23681105.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GTNH
            Center clone name: CH230-23?H8
            ----- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 227037 bases at least Q40
            Consensus quality: 229537 bases at least Q30
            Consensus quality: 23125 bases at least Q20
            Consensus quality: 235071; sum-of-contigs estimation
            Estimated insert size: 235071; sum-of-contigs estimation
            Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            *
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 2 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            *   This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            *
            *   108068: contig of 108068 bp in length
            *   108069   108068: gap of unknown length
            *   108169   235102: contig of 126934 bp in length.
FEATURES             Location/Qualifiers
     source          1..235102
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-23?H8"
     misc_feature    1..1143
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(3003..3706)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ101611"
     misc_feature    108169..110729
                     /note="wgs_contig"
ORIGIN
```

```
RESULT 4
AC110369/c
LOCUS       AC110369              258368 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-49K10, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC110369
VERSION     AC110369.4  GI:23195246
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258368)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,B., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,H., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,D., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garcia,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,V., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,J.,
            Hollins,B., Howells,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindashtne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,R., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,R., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwokelemeh,O., Okwuonu,G., Olairunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
```

```
Db    89327  CTTGTTACAGGGCCACAGCTAAAAGAAAGCAGGTCCAGCTGGA--CCCAGTTGTGTGCA  89270
             ||||   ||| ||| |||   ||||||||| |||  |||||||||   |||||||||||||
Qy      172  GCCCCAGCCCAGTTGCTCATTGCGGGGCTCCGGGAGCCGCACGAGCGAGGCTGAGCAGATG  231
             ||||||||||    ||  |||||||||||||  |||||||||||  ||||||||||||||||
Db    89269  GCTCTCAGTCCTGTACTCATTTAGACTCCAGGGACCTCTCATAAGCACTGCCAAGCAGT  89210
             ||||  |||| ||||| |||  |||||||  |||||| ||||| |||||||||||||||||
Qy      232  TGTTCCAGATGGTTGGGGAACTGGAGAGGGCCCGGCACAGGCGCCCGTGCAGGGAACCCGAGG  291
             ||||||||||   ||||  ||||| |||||||| ||||||| |||||| |||||||
Db    89209  GTGGTGCCAGACGACCCGGGACGTCTCAGTACAGGCTCATGT------------  89164
             |||||||||||||||||||||| |||||||| |||||||||||
Qy      292  GCTGTAGGCCCCGTGCCACTGCATGCCTCAGGCCCTGGTGGTCCGCCCACCACACCCCTA  351
             |||||||||||||||||  ||||||||||||  ||||||||||||||||||  ||||||||
Db    89163  ------------CACCCACATGCCTCAGGCCCTGGTGGTCCGCCCACCACACCCCTG  89119
             
Qy      352  CTGCTGACGGCAGCAGAAATCTGAGCCCG---GGAAGGGGTCCAGGGAAGTTCGTGAACCA  408
             ||||||||||||  ||||||||||||||||   ||  |||||| |||||||||||||||||
Db    89118  TTGCTGATGGTGGCAGAAATCTGAGCCAGTGATGAAGGTTTGGGGGGACGTTCATTGTGCTG  89059
             |||||||  |||||||||||||||||| ||  |||||| |||||||||||||||||||| |
Qy      409  TCTAGCAAGTCGGGCGTGGGGTCTGGCCAAGTTAGACAGCAG  448
             |||||||||||||||||||| |||  |||||||||||||||
Db    89058  CCTAGCAAGTAGGGCCTGGGGATATGGCCAAGTTGGGACAAAG  89019
```

```
  AUTHORS   Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snajs,D.,
            Sneed,A., Sodergren,E., Song,X-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Vales,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,L., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Holt,R.A., Smith,D.R., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 258368)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 258368)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2002 this sequence version replaced gi:21744398.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads. Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            --------------------------------------------------------------
            Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            --------------------------------------------------------------
            Project Information
              Center project name: GRPY
              Center clone name: CH230-49K10
            --------------------------------------------------------------
            Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 236348 bases at least Q40
              Consensus quality: 238877 bases at least Q30
              Consensus quality: 240713 bases at least Q20
              Estimated insert size: 259061; sum-of-contigs estimation
              Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------------------------------------------------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 3 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            --------------------------------------------------------------
FEATURES             Location/Qualifiers
     source          1..258368
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
```

```
     misc_feature    /clone="CH230-49K10"
                     1..1248
                     /note="wgs_end_extension"
                     /clone_end="T7"
     misc_feature    6211..7032
                     /note="clone_boundary"
                     /clone_end="T7"
                     /note="clone_boundary
                     /site:EcoRI
                     end_sequence:BH282841"
     misc_feature    78223..79518
                     /note="wgs_contig"
     misc_feature    102260..104120
                     /note="wgs_contig"

ORIGIN
```

```
Query Match        7.4%;    Score 77.2;   DB 2;   Length 258368;
Best Local Similarity   58.0%;   Pred. No. 2.6e-06;
Matches 232;   Conservative  0;   Mismatches 133;   Indels  35;   Gaps  4;

Qy      53 GAGGGCATATTCAGTCCCCATTTTTCAGATGAG-GAGTTGAGCGCCCAGAGGAACCTAAGTAA 111
           |||||  ||||||||| || ||||||||||||| ||||||||||| ||||| ||| |||||
Db   82321 GGGGTGTCTTCAGTCCCCACTTGTTACAGATGGGGAAAGTCAAAGGCCCAGAGAGGGTAACAGA 82262

Qy     112 TCTGTCTGAGGGCCACACAGCTAGAAAAGCAGCAGCCCAGCCGAACCCCTGGTGTGTGTGCA 171
Db   82261 CTTGTTACAGGCCACACAGCTAAAAGAAAGCAGGTCCAGCTGGA--CCCAGTTGTGTGTGTCA 82204

Qy     172 GCCCCCAGCCCACGTTGCTCATTGCGGGGCCCTCGGGAGCCCACGACGGAGCGTGACGCATG 231
Db   82203 GCTCTCAGTCCCTGTACTCATTTTAGACTCCCCAGGGACCCCTCATCATAAGCACTGCCAAGCAGT 82144

Qy     232 TGTTCCAGATGGTGGGGAACTGAGAGAGACCCCGGCCACAGGCCCCGTGCAGGGGAACCCCGAGG 291
Db   82143 GTGGGTGCCAGAGACCGGGAGCCTGGAGGTCTCAGTACACGGCTCATGT------ 82098

Qy     292 GCTGTTAGGCCCCGTGCCACTGCAATGCCTCAGGCCCTGTGGTCCTGGCCAGCCAGCCCCTA 351
Db       --CACCCACATGCCTCAGGCCTGTCGTTCCTGGCCAAGCCACCAGCCCCTG 82053

Qy     352 CTGCTGACGGGCACCAGGAAATCTGAGCCCG---GGAAGGGGTCAGGGGAAGTTCGTGAACCA 408
Db   82052 TTGCTGATGGTGGCGAGAAATCTGAGCCCAGTGATGAAGGGTTTGGGGGACCTTCAATGTGCTG 81993

Qy     409 TCTAGCAAGTCGGGGCCTGGGGGTCGGGCCCAAGTTAGACACAG 448
Db   81992 CCTAGCAAGTAGGCCCTGGGATATGGCCAAGTTGGACAAAG 81953
```

```
RESULT 5
LOCUS       AC134075
DEFINITION  Rattus norvegicus clone CH230-2H18, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC134075
VERSION     AC134075.2  GI:24941374
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
LOCUS       AC134075           264075 bp    DNA    linear   HTG 13-NOV-2002
REFERENCE   1  (bases 1 to 264075)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Albrecks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleland,C.D., Cox,C.C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
```

```
          *  (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
          *  NOTE: This is a 'working draft' sequence. It currently
          *  consists of 2 contigs. The true order of the pieces
          *  is not known and their order in this sequence record is
          *  arbitrary. Gaps between the contigs are represented as
          *  runs of N, but the exact sizes of the gaps are unknown.
          *  This record will be updated with the finished sequence
          *  as soon as it is available and the accession number will
          *  be preserved.
          *
          *  1      262529: contig of 262529 bp in length
          *  262530  262629: gap of unknown length
          *  262630  264075: contig of 1446 bp in length.

FEATURES             Location/Qualifiers
     source          1..264075
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-2H18"
     misc_feature    1..1475
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1858..153329
                     /note="clone_boundary
                     clone_end:T7"
                     site:
                     end_sequence:BH287375"
     misc_feature    258169..258875
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH287377"
     misc_feature    259534..262529
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN
```

```
REFERENCE   (bases 1 to 264075)
AUTHORS   Barnhart,C., Edgar,D., Edwards,C.C., Blhaj,C., Escotto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,R., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.L., Lewis,L.,
          Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
          Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newton,J., Newtson,J., Nguyen,N., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
          Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
          Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
          Sutton,A., Svetek,T., Tabor,P., Tamerisa,K., Tang,H.,
          Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
          Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 2  (bases 1 to 264075)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 264075)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 13, 2002 this sequence version replaced gi:23267440.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contig within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -----------  Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -----------  Project Information
          Center project name: TUQB
          Center clone name: CH230-2H18
          -----------  Summary Statistics
          Assembly program: Phrap, version 0.990329
          Consensus quality: 225536 bases at least Q40
          Consensus quality: 228408 bases at least Q30
          Consensus quality: 230612 bases at least Q20
          Estimated insert size: 229980; sum-of-contigs estimation
          Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          -----------
          * NOTE: Estimated insert size may differ from sequence length
```

```
RESULT 6
AF429315
LOCUS       AF429315      125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
```

```
Query Match   58.0%;  Score 77.2;  DB 2;  Length 264075;
Best Local Similarity  58.0%;
Matches 232;  Conservative  0;  Mismatches 133;  Indels 35;  Gaps 4;
               7.4%;  Pred. No. 2.5e-06;

Qy     53 GAGGGGCATATTCAGTCCCATTTTTCAGATGAG-GAGTTGAGGCCCAGAGAACGTAAGTAA 111
           |||||||  |||||||| ||  |||||||  |||| |||| |||||||||||||||  ||
Db  88478 GGGGTCCTCTTCAGTCCCACTTGTTTAGATGGGAAAGTCAAGGCCCAGAGAGGGTAACAGA 88537

Qy    112 TCTGTGTCTGAGGCCCACACAGCTAGAAAAGCAGCCAGCCCAGCCGGAAGCCCTGTGTGCA 171
           | | | || |||||||||||||||| |||||||| |  | |||| ||   |||||||||||
Db  88538 CTTGTTACAGGGCCACACAGCTAAAGAAAGCAGAGGTCCAGGTGGA--CCCAGTTGTGTGCA 88595

Qy    172 GCCCCCAGCCCCAGTTGCTCATTGCGGGGGCGTCGGGAGCCCACGAGCGGAGGCTGAGCAGCATG 231
           ||  |  ||||  |||| |||||||   |||  |   || |||| |   |||  |||||||
Db  88596 GCTCCTCAGTCCTGTACTCATTTTAGACTCCAGGGACCTCTCATAACCACTCCCAAGCAGT 88655

Qy    232 TGTTCCAGATGGTGGGGAACTCGGAGAGAGAGCCCGGACACAGCGCCCGTGCAGGGAACCCCGAGG 291
           ||  | |||   | ||   | |||| ||| ||||  |||||  |||| || |||||||||
Db  88656 GTGGGTGCCAGAGACCCGGACCTGGAGGTCTCAGTACAGGCTCATGT------------ 88701

Qy    292 GCTGTGTAGGGCCCGTGCCACTCGCATGCCTCAGGCCCTGTGGTCCTGGCAGCCACAGCCCCTA 351
           ||||  |  ||| | ||||||| ||||||||||||||  ||| |||| | |||||||||||||
Db  88702 ----------CACCCACATGCCTCAGGCCGTGGCGTCCTCAGGGAAGCCACAGCCCCTG 88746

Qy    352 CTGCTGACACGGCGCAGCAGGAATCTGAGCCCG---GGAAGGGGTCCAGGGAGAAGTTCGTGAACCA 408
           |||||||  | | |||  | | |||||||||     || || | |||| | |||| |||||||| ||
Db  88747 TTGCTGATGGTGGCCACAAAATCTGAGCCAGTGATGAAGGTTTGGGGGGACGTTCATGTGGCTG 88806

Qy    409 TCTAGCAAAGTCGGGGTCGGGGGTGGGGGCCAAGTTAGACACAG 448
           |||||||| ||  ||| || || |||||| ||||||||| |  ||
Db  88807 CCTAGCAAGTAGGCCTGGGATATGGCCCAAGTTGGACAAAG 88846
```

```
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
   MEDLINE  21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1..125020
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-like 2 (HDL2)"
     repeat_region   complement(35581..35746)
                     /rpt_type=tandem
                     /rpt_unit="ctg"
     gene            complement(<36507..>36887)
                     /gene="JPH3"
                     /note="synonym: JP3"
     mRNA            complement(<36507..>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
                     /complement(<36507..36887)
     CDS             complement(<36507..36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRENFDDGSYCCGWEDGKAHGHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNYYQGTWAQGKRHGIGIGLESKGKWVYKGEWTHGFGKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
ORIGIN
```

```
Query Match        5.4%;  Score 56.4;  DB 9;  Length 125020;
Best Local Similarity  10.8%;  Pred. No. 0.058;
Matches  95;  Conservative  398;  Mismatches  376;  Indels  14;  Gaps  4;

Qy    127 ACAGCTAGAAAGACACCAGGCCCAGGCGCGAACCCTGGTGTGTGTGCAGCCCCCCAGTT 186
          ::: ::  ::::    :: :: :: ::  ::: ::: : : : :::::: :  ::
Db  16895 MSMBMSVSYSVKMHSMHASBSCMHWBKCMTWSCCMSMMYKSSWWGSSWGMCCWGRRRSKG 16954

Qy    187 GCTCATTGCGGGGCTCGGGAGCCACGAGCGGTCTGAGCAGCATGTGTTCCAGATGGTGG 246
          :: ::: :::  :::::::   :::::    :::  :::::::::::: ::::::: ::
Db  16955 WKWYSRGMSRSKSMRTGGSKMRSSMMCTSSCYASMCCMCCWSCCCMRSCCCMCRSYCC 17014

Qy    247 GAACTGGAGAGAGCCCGGGAACCCGGTCGCAGGAGAAAGCTGCCGAGGGCGGCCCCGTG 306
          ::::      :::: :::: :  ::  :  :::: ::: :::: ::  ::::::
Db  17015 CMRYCCACKCYMSSSYWTMSASYWKRSMSMKCWSRMCSRSRSSKCKSRGCGSMGKG 17074

Qy    307 CCACTGCCATGCCCTCAGGCCCTGTGGTCCTCTGGACGGCACGAGGCGGCAGCA 366
          ::  ::::  ::: ::::: ::  ::: ::::::   :::: ::::::: ::
Db  17075 KGGKSVGRKTKRKSRGKMKGAKWMYMYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCM 17134

Qy    367 GGAATCTGAGCCCGGGGAGGGTTCAGGGAAGTT--CGTGAACCACATCAGTCAAGTCGGGGCT 424
          :::::::::: :::: ::::::  :::::::::     ::::: :::: :::::::::::::
Db  17135 TSRSAMCSYCYAKCKSMCYSCYYGMSMKGYYYACSYRGSMSSKYCMRGSYTSYSCGGCC 17194

Qy    425 GGGGGTGTGGCCAAGGTTAGACACAGAGATGCTAGGCGCCCGTGTGGACTCAGAAATTGGCAGCTCT 484
          :::::::::::::  ::  : : :::  ::: ::::::: :: ::::::     :::::: ::  :
Db  17195 TTTTTCCCCCNANTGGGGAGAAGCTTTTNCNKKYTSYPRKRNNGCAMCKXYNNYNNSWRSSCR 17254

Qy    485 TTTGGCCCAGAGGGGGCCACGCTGTGTCCGGGCGTCTTGGCTGCTTCAGATGGTGGG 544
          ::::    :::: :::: ::: : : ::::: ::  ::  ::  ::::::: ::::
Db  17255 AGMSCTKYKSSMTMSMASYCWCMSMYCSMRSNASRGMWSSYKMWKSSWRMSCYMKCC 17314

Qy    545 GTCTTCCACTACA----CCCCGCCTGGACACTGCTGTTAGGCCCCAGGGGACC 600
          :::: ::: : :    :  :::: ::: : ::: :::  ::::::::::::
Db  17315 WCMKMCYCMMRSMRSRGMSYYMYASWKSSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRM 17374

Qy    601 AGCCTGGAGCCCATGAGGAGAAGGAGGGCCAGTTGTTCTCCTGTAAGGGTATTGCTGTAGCATGAG 660
          :: ::::: :::::::: ::: :::::::: :::::::::: :::::::::: ::::  ::::
Db  17375 KGMGKWSRCWGSMWGSGASRSSCYKCYKSMRCSMWSSKCYRCAGCMMMKGGYMRYMRCWSMK 17434

Qy    661 GGAACAGACRAAGGCCCCAGGGGGGACTRAACCCGAGATCCACCCCCGGCCTCACTCCCGTGTG 720
          :::: :::::::::::::   ::::::   ::::::::  ::::::: : :::: :::: :
Db  17435 RRWWGKSAMYRMRRMWKGRGAMMMCMKCYSRMSRCMWMMKSYCASCRSCAMMSGTYMKCA 17494

Qy    721 GCTCACGGCGAATATCCTAACCTCTCTCTCAGAGCCTCCTGCCCAGCCTAGCAGGGGGTTCCAGTG 780
          :: ::::::  :::::::: ::: :::::: ::::::::::: :: :::::::::::::
Db  17495 SYMCMSSYWCTCWCMTSYMSYYRCTCWCKGWSSYTKSKSWSSSSMSSYKGRRKSYSMCCTS 17554

Qy    781 AG------GGGGTGAGGAAGCCCAGCACGTGGAAGCCCTTTTTTAACCATTCTCGGGGGTGA 834
          ::      :::::: ::::::::::::: ::::::::::    :::::::: ::::::::
Db  17555 RGAMSCWRRCCYMRGASSMRAGSMSRRRAKGRSWGGRSKWWMTGGMWRSKXYYYCYTGRRM 17614

Qy    835 GCCAGCCCCTTCCCAAAATGCCTGGTGTCCTGTGTCACTGCACTGCTGTGTGTGGGGGTTCCCAAC 894
          :::::::::: :::::  ::::::::::::::::  :::::::::::::::::   ::::::
Db  17615 MMTYMCCCWRRRSYMYRSMSAMGMRKSSWSGMRMMGSASSRRCKSASRSSWCSRRMKGM 17674

Qy    895 GGGGCTCAGTGTGTGGGCTGAGGTGTGGCGTCTGAACTGGGG--ACAGGGGTTCTCAGGAGAGGCCT 952
          ::::::::::::::::: ::::::::::::::::::     ::  ::::::::::: :::::::::
Db  17675 RGGSCWSSKOWWGGSRSRSASSKCKGSRGMRRRSKRSSKYRKGRGKKRSMTKSKGSKGKSK 17734

Qy    953 CCTCCCTCCTTGCCCCACTGGGGCATAGGGCCTTCTGGGAGCTGGCAGC 995
          :::  :::::::: :::::::::: ::::::::::::::::::::::
Db  17735 CWKRSGSMTSSCYYYSASSCMWWMSSKSCMCCCMWMMRRCACC 17777
```

```
RESULT 7
AC021197
LOCUS       AC021197      172105 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 19 clone RP11-798B19, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC021197
VERSION     AC021197.5  GI:8570404
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172105)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172105)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 17, 2000 this sequence version replaced gi:7235351.
            --------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            --------- Project Information -----------
```

```
Center project name: H_NH0798B19

--------- Summary Statistics ---------

Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET: 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160434 bases at least Q40
Consensus quality: 16435 bases at least Q30
Consensus quality: 16763 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 170105; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 19395..23283 bases; sum-of-contigs
Quality coverage: 4.02 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 21 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

     1    1301: contig of 1301 bp in length
  1302    1401: gap of unknown length
  1402    4060: contig of 2659 bp in length
  4061    4160: gap of unknown length
  4161    7024: contig of 2864 bp in length
  7025    7124: gap of unknown length
  7125    9662: contig of 2538 bp in length
  9663    9762: gap of unknown length
  9763   13070: contig of 3308 bp in length
 13071   13170: gap of unknown length
 13171   15928: contig of 2758 bp in length
 15929   16028: gap of unknown length
 16029   19294: contig of 3266 bp in length
 19295   19394: gap of unknown length
 19395   23283: contig of 3889 bp in length
 23284   23383: gap of unknown length
 23384   27272: contig of 3889 bp in length
 27273   27372: gap of unknown length
 27373   32535: contig of 5163 bp in length
 32536   32635: gap of unknown length
 32636   40106: contig of 7471 bp in length
 40107   40206: gap of unknown length
 40207   45656: contig of 5450 bp in length
 45657   45756: gap of unknown length
 45757   52232: contig of 6476 bp in length
 52233   52332: gap of unknown length
 52333   59204: contig of 6872 bp in length
 59205   59304: gap of unknown length
 59305   68409: contig of 9105 bp in length
 68410   68509: gap of unknown length
 68510   76838: contig of 8329 bp in length
 76839   76938: gap of unknown length
 76939   88902: contig of 11964 bp in length
 88903   89002: gap of unknown length
 89003  100235: contig of 11233 bp in length
100236  100335: gap of unknown length
100336  117760: contig of 17425 bp in length
117761  117860: gap of unknown length
117861  142316: contig of 24456 bp in length
142317  142416: gap of unknown length
142417  172105: contig of 29689 bp in length

FEATURES             Location/Qualifiers
     source          1..172105
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-798B19"
     misc_feature    1..1301
                     /note="assembly_name:Contig11"
     misc_feature    1402..4060
                     /note="assembly_name:Contig12"
     misc_feature    4161..7024
                     /note="assembly_name:Contig13"
     misc_feature    7125..9662
                     /note="assembly_name:Contig14"
     misc_feature    9763..13070
                     /note="assembly_name:Contig15"
     misc_feature    13171..15928
                     /note="assembly_name:Contig16"
     misc_feature    16029..19294
                     /note="assembly_name:Contig17"
     misc_feature    19395..23283
                     /note="assembly_name:Contig18"
     misc_feature    23384..27272
                     /note="assembly_name:Contig19"
     misc_feature    27373..32535
                     clone_end:SP6
                     vector_side:right"
                     /note="assembly_name:Contig20"
     misc_feature    32636..40106
                     /note="assembly_name:Contig21"
     misc_feature    40207..45656
                     /note="assembly_name:Contig22"
     misc_feature    45757..52232
                     /note="assembly_name:Contig23"
     misc_feature    52333..59204
                     /note="assembly_name:Contig24"
     misc_feature    59305..68409
                     /note="assembly_name:Contig25"
     misc_feature    68510..76838
                     /note="assembly_name:Contig26"
     misc_feature    76939..88902
                     /note="assembly_name:Contig27"
     misc_feature    89003..100235
                     /note="assembly_name:Contig28"
     misc_feature    100336..117760
                     /note="assembly_name:Contig29"
     misc_feature    117861..142316
                     /note="assembly_name:Contig30"
     misc_feature    142417..172105
                     /note="assembly_name:Contig31"
ORIGIN
```

```
RESULT 8
AP005019
LOCUS       AP005019               98501 bp    DNA     linear   PRI 30-MAR-2002
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-169D4, complete
            sequences.
ACCESSION   AP005019
VERSION     AP005019.1  GI:19979835
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match        5.4%;  Score 56.2;  DB 2;  Length 172105;
Best Local Similarity  65.6%;
Matches  82;  Conservative   0;  Mismatches  43;  Indels  0;  Gaps  0;

Qy      55 GGGCATATTCAGTCCCCATTTTTCAGATGAGGAGTTGAGGCGCCCAGAGAAACGTAAGTAATCT 114
           |||||  |||||||||| |||||| |||| |||||  ||||||||||| |||||||  |||
Db  103864 GGCTGTAGTTATTCCCATTTTACAGAAGAGGGGGTTGAGGCCCAGAGAGGGGTGAGGAAGTT 103923

Qy     115 GTCTGAGGGCCACACAGCTAGAAAAGCAGCAGCCAGCGACCCCTTGGTGTGTTGTCAGCC 174
           |||||||| ||| |||| ||||||||||| ||||||||||||||||| |||| ||||
Db  103924 GCCTGAAGCCACACAGCAGCCACGGCCCTATCAGAATCTCCACCCCGTGCACCT 103983

Qy     175 CCCAG 179
           |||||
Db  103984 CTCAG 103988
```

```
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2002)
REFERENCE   2 (bases 1 to 98501)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
FEATURES             Location/Qualifiers
     source          1..98501
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-169D4"
ORIGIN

Query Match     5.4%;  Score 55.8;  DB 9;  Length 98501;
Best Local Similarity   70.1%;  Pred. No. 0.083;
Matches  89;  Conservative   0;  Mismatches  37;  Indels   1;  Gaps   1;

Qy     67 TCCCATTTTTCAGATGAGGAGTTGAGGGCCCAGAGAGAACGTAAGTAATCTGTCTGAGGCTCAC 126
          |||| |||||||| || || ||||||||||||||| ||| ||| |||| | | ||||| ||||
Db  77268 TCCCCATTTTTCAGGTGTGAAACTGAGGGCCCAGAGGACGGTTTAGTGACTTGCTCAAGGTCAC 77327

Qy    127 ACAGCTAGAAAAGCAGCCAGGCC-CAGCCGAACCCCTGGTGTGTGTTCGCAGCCCCCAGCCCAGT 185
          ||||||  |||||||| |   || |||  |||||||| | ||| ||  |||| |||| |||||
Db  77328 ACAGCTCAAAAGCAGCAAAAGCCGAAGCTGAAACCCCAGGGTCTGTCCAGCTCCCTGGCCAGG 77387

Qy    186 TGCTCAT 192
          ||
Db  77388 CCATCTT 77394


RESULT 9
LOCUS       AR368084              111282 bp    DNA     linear   PAT 12-SEP-2003
DEFINITION  Sequence 3 from patent US 6376225.
ACCESSION   AR368084
VERSION     AR368084.1  GI:34601657
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 111282)
  AUTHORS   Wei,M.-H., Wang,X., Merkulov,G.V., Di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human phosphodiesterase proteins, nucleic acid molecules
            encoding human phosphodiesterase proteins, and uses thereof
  JOURNAL   Patent: US 6376225-A 3 23-APR-2002;
FEATURES             Location/Qualifiers
     source          1..111282
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Query Match     5.4%;  Score 55.8;  DB 6;  Length 111282;
Best Local Similarity   70.1%;  Pred. No. 0.08;
Matches  89;  Conservative   0;  Mismatches  37;  Indels   1;  Gaps   1;

Qy     67 TCCCATTTTTCAGATGAGGAGTTGAGGGCCCAGAGAGAACGTAAGTAATCTGTCTGAGGCTCAC 126
          |||| |||||||| || || ||||||||||||||| ||| ||| |||| | | ||||| ||||
Db  93162 TCCCCATTTTTCAGGTGTGAAACTGAGGGCCCAGAGGACGGTTTAGTGACTTGCTCAAGGTCAC 93103

Qy    127 ACAGCTAGAAAAGCAGCCAGGCC-CAGCCGAACCCCTGGTGTGTGTTCGCAGCCCCCAGCCCAGT 185
          ||||||  |||||||| |   || |||  |||||||| | ||| ||  |||| |||| |||||
Db  93102 ACAGCTCAAAAGCAGCAAAAGCCGAAGCTGAAACCCCAGGGTCTGTCCAGCTCCCTGGCCAGG 93043

Qy    186 TGCTCAT 192
          ||
Db  93042 CCATCTT 93036


RESULT 10
LOCUS       AX662038/c
            AX662038              111282 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 3 from Patent WO02053718.
ACCESSION   AX662038
VERSION     AX662038.1  GI:29163011
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wei,M.H., Wang,X., Merkulov,G., di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human phosphodiesterase proteins, nucleic acid molecules
            encoding human phosphodiesterase proteins, and uses thereof
  JOURNAL   Patent: WO 02053718-A 3 11-JUL-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..111282
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match     5.4%;  Score 55.8;  DB 6;  Length 111282;
Best Local Similarity   70.1%;  Pred. No. 0.08;
Matches  89;  Conservative   0;  Mismatches  37;  Indels   1;  Gaps   1;

Qy     67 TCCCATTTTTCAGATGAGGAGTTGAGGGCCCAGAGAGAACGTAAGTAATCTGTCTGAGGCTCAC 126
          |||| |||||||| || || ||||||||||||||| ||| ||| |||| | | ||||| ||||
Db  93162 TCCCCATTTTTCAGGTGTGAAACTGAGGGCCCAGAGGACGGTTTAGTGACTTGCTCAAGGTCAC 93103

Qy    127 ACAGCTAGAAAAGCAGCCAGGCC-CAGCCGAACCCCTGGTGTGTGTTCGCAGCCCCCAGCCCAGT 185
          ||||||  |||||||| |   || |||  |||||||| | ||| ||  |||| |||| |||||
Db  93102 ACAGCTCAAAAGCAGCAAAAGCCGAAGCTGAAACCCCAGGGTCTGTCCAGCTCCCTGGCCAGG 93043

Qy    186 TGCTCAT 192
          ||
Db  93042 CCATCTT 93036


RESULT 11
LOCUS       AC026173
            AC026173              194719 bp    DNA     linear   HTG 05-APR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-169D4 map 3p, *** SEQUENCING
            IN PROGRESS ***, 62 unordered pieces.
ACCESSION   AC026173
VERSION     AC026173.3  GI:20042926
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 194719)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
            Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
            Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished

Qy    186 TGCTCAT 192
          ||
Db  93042 CCATCTT 93036
```

```
REFERENCE   2  (bases 1 to 194719)
AUTHORS     Chen,J., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B.,
            Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,
            Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,
            Li,G., Li,J., Li,L., Bao,Q., Wang,X., Song,L., Zhang,L., Guo,D.,
            Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.
            and Yang,H.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAR-2000) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   3  (bases 1 to 194719)
AUTHORS     Bao,J., Bao,Q., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,T., Liu,J., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,L., Yang,L., Zhang,M.,
            Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
TITLE       Direct Submission
JOURNAL     Submitted (05-APR-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     On Apr 5, 2002 this sequence version replaced gi:8101143.
            ---------------Genome Center
            Center:Beijing Center
            Center project name:1% project
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            ---------------Project Information
            Center project name:1% project
            Center clone name: RP11-169D4
            ---------------Summary Statistics
            Sequencing vector:pUC18; 100% of reads
            Chemistry:Dye-terminator: ET 55% of reads
            Chemistry:Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 37507 bases at least Q40
            Consensus quality: 45201 bases at least Q30
            Consensus quality: 51060 bases at least Q20
            Insert size: 52554; sum-of-contigs
            Quality coverage: 1.65x in Q20 bases;sum-of-contigs
            ---------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 62 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *        1       1336: contig of 1336 bp in length
          *     1337       1436: gap of unknown length
          *     1437       2122: contig of 686 bp in length
          *     2123       2222: gap of unknown length
          *     2223       2945: contig of 723 bp in length
          *     2946       3045: gap of unknown length
          *     3046       3860: contig of 815 bp in length
          *     3861       3960: gap of unknown length
          *     3961       4756: contig of 796 bp in length
          *     4757       4856: gap of unknown length
          *     4857       5429: contig of 573 bp in length
          *     5430       5529: gap of unknown length
          *     5530       6114: contig of 585 bp in length
          *     6115       6214: gap of unknown length
          *     6215       7064: contig of 850 bp in length
          *     7065       7164: gap of unknown length
          *     7165       7982: contig of 818 bp in length

          *     7983       8082: gap of unknown length
          *     8083       8795: contig of 713 bp in length
          *     8796       8895: gap of unknown length
          *     8896       9588: contig of 693 bp in length
          *     9589       9688: gap of unknown length
          *     9689      10512: contig of 824 bp in length
          *    10513      10612: gap of unknown length
          *    10613      11423: contig of 811 bp in length
          *    11424      11523: gap of unknown length
          *    11524      12508: contig of 985 bp in length
          *    12509      12608: gap of unknown length
          *    12609      13724: contig of 1116 bp in length
          *    13725      13824: gap of unknown length
          *    13825      14905: contig of 1081 bp in length
          *    14906      15005: gap of unknown length
          *    15006      16240: contig of 1235 bp in length
          *    16241      16340: gap of unknown length
          *    16341      17952: contig of 1612 bp in length
          *    17953      18052: gap of unknown length
          *    18053      19287: contig of 1235 bp in length
          *    19288      19387: gap of unknown length
          *    19388      20388: contig of 1001 bp in length
          *    20389      20488: gap of unknown length
          *    20489      21703: contig of 1215 bp in length
          *    21704      21803: gap of unknown length
          *    21804      22654: contig of 851 bp in length
          *    22655      22754: gap of unknown length
          *    22755      24998: contig of 2244 bp in length
          *    24999      25098: gap of unknown length
          *    25099      26465: contig of 1367 bp in length
          *    26466      26565: gap of unknown length
          *    26566      28332: contig of 1767 bp in length
          *    28333      28432: gap of unknown length
          *    28433      29906: contig of 1474 bp in length
          *    29907      30006: gap of unknown length
          *    30007      32797: contig of 2791 bp in length
          *    32798      32897: gap of unknown length
          *    32898      35022: contig of 2125 bp in length
          *    35023      35122: gap of unknown length
          *    35123      37579: contig of 2457 bp in length
          *    37580      37679: gap of unknown length
          *    37680      39880: contig of 2201 bp in length
          *    39881      39980: gap of unknown length
          *    39981      42263: contig of 2283 bp in length
          *    42264      42363: gap of unknown length
          *    42364      44571: contig of 2208 bp in length
          *    44572      44671: gap of unknown length
          *    44672      48146: contig of 3475 bp in length
          *    48147      48246: gap of unknown length
          *    48247      50501: contig of 2255 bp in length
          *    50502      50601: gap of unknown length
          *    50602      52997: contig of 2396 bp in length
          *    52998      53097: gap of unknown length
          *    53098      56547: contig of 3450 bp in length
          *    56548      56647: gap of unknown length
          *    56648      60181: contig of 3534 bp in length
          *    60182      60281: gap of unknown length
          *    60282      63280: contig of 2899 bp in length
          *    63281      63380: gap of unknown length
          *    63381      66358: contig of 3078 bp in length
          *    66359      66458: gap of unknown length
          *    66459      67956: contig of 1498 bp in length
          *    67957      68056: gap of unknown length
          *    68057      71612: contig of 3556 bp in length
          *    71613      71712: gap of unknown length
          *    71713      74532: contig of 2820 bp in length
          *    74533      74632: gap of unknown length
          *    74633      78476: contig of 3844 bp in length
          *    78477      78576: gap of unknown length
          *    78577      82244: contig of 3668 bp in length
          *    82245      82344: gap of unknown length
          *    82345      86135: contig of 3791 bp in length
          *    86136      86235: gap of unknown length
```

RESULT 12

Qy   186 TGCTCAT 192
         | | | | | | |
Db 44910 CCATCTT 44916

Qy   127 ACAGCTAGAAAAGCAGCCAGCGCC--CAGCGCAAACCCTCGGTGTGTGCAGCCCCCAGCCCAGT 185
         | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44850 ACAGCTCAAAAAGCACCAAAACCCGAAGCTGAAACCAGGTCTGTCGTCCCTGSCCAGG 44909

Qy   67 TCCCATTTTTCAGATGAGAGGAGTTGAGGCGCCCAGAGAGGAACGTTAAGTAAATCTGTCTGAGGCCAC 126
         | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44790 TCCCCATTTTCAGGTGTGAAAACTGAGGCCCCAGAGACGCGTTTAGTGACTTGCTCAAGGTCAC 44849

Query Match   5.4%;   Score 55.8;   DB 2; Length 194719;
Best Local Similarity   70.1%;   Pred. No. 0.067; DB 37; Indels 1; Gaps 1;
Matches   89; Conservative   0; Mismatches   37;   Indels 1; Gaps 1;

*   86236   90642: contig of 4407 bp in length
*   90643   90742: gap of unknown length
*   90743   95083: contig of 4341 bp in length
*   95084   95183: gap of unknown length
*   95184   98380: contig of 3197 bp in length
*   98381   98480: gap of unknown length
*   98481   102414: contig of 3934 bp in length
*   102415   102514: gap of unknown length
*   102515   106775: contig of 4261 bp in length
*   106776   106875: gap of unknown length
*   106876   112291: contig of 5416 bp in length
*   112292   112391: gap of unknown length
*   112392   116740: contig of 4349 bp in length
*   116741   116840: gap of unknown length
*   116841   121741: contig of 4901 bp in length
*   121742   121841: gap of unknown length
*   121842   127122: contig of 5281 bp in length
*   127123   127222: gap of unknown length
*   127223   132291: contig of 5069 bp in length
*   132292   132391: gap of unknown length
*   132392   139100: contig of 6709 bp in length
*   139101   139200: gap of unknown length
*   139201   148094: contig of 8894 bp in length
*   148095   148194: gap of unknown length
*   148195   155530: contig of 7336 bp in length
*   155531   155630: gap of unknown length
*   155631   163081: contig of 7451 bp in length
*   163082   163181: gap of unknown length
*   163182   171479: contig of 8298 bp in length
*   171480   171579: gap of unknown length

Db 44850 Homo sapiens chromosome 11 clone RP11-796A3 map 11, WORKING DRAFT

Qy   67 Homo sapiens chromosome 11, clone RP11-796A3

Db 44790 Homo sapiens (human)

LOCUS   AC055829   197022 bp   DNA   linear   HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 11 clone RP11-796A3 map 11, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC055829
VERSION   AC055829.3 GI:9966988
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 197022)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Unpublished
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 197022)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Cheepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,H.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J.T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,D.C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rochman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 197022)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camrata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,H., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Tesfaye,S.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Sep 5, 2000 this sequence version replaced gi:7798805.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L8652
            Center clone name: 796_A_3
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 96% of reads
            Sequencing vector: Plasmid; n/a; %-0.f%% of reads
            4.2305334150827?Chemistry: Dye-terminator Big Dye; 100% of
            reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 183643 bases at least Q40
            Consensus quality: 189993 bases at least Q30
            Consensus quality: 192582 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 194322; sum-of-contigs
            Quality coverage: 4.5 in Q20 bases; agarose-fp
            Quality cov.
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will

* be preserved.

```
FEATURES             Location/Qualifiers
     source          1..197022
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-796A3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..11810
                     /note="assembly_fragment
                     vector_side:left"
                     clone_end:SP6
     misc_feature    1911..13354
                     /note="assembly_fragment"
     misc_feature    1455..15818
                     /note="assembly_fragment"
```

```
1:       contig of 11810 bp in length
11811:   gap of 100 bp
11910:   contig of 1444 bp in length
13354:   gap of 100 bp
13455:   contig of 2364 bp in length
15818:   gap of 100 bp
15819:   contig of 2519 bp in length
18437:   gap of 100 bp
18538:   contig of 2138 bp in length
20675:   gap of 100 bp
20776:   contig of 1279 bp in length
22054:   gap of 100 bp
22155:   contig of 3261 bp in length
25415:   gap of 100 bp
25516:   contig of 2365 bp in length
27880:   gap of 100 bp
27981:   contig of 2879 bp in length
30859:   gap of 100 bp
30960:   contig of 4527 bp in length
35486:   gap of 100 bp
35587:   contig of 2728 bp in length
38314:   gap of 100 bp
38415:   contig of 3660 bp in length
42074:   gap of 100 bp
42175:   contig of 3738 bp in length
45912:   gap of 100 bp
46013:   contig of 5592 bp in length
51604:   gap of 100 bp
51705:   contig of 5681 bp in length
57385:   gap of 100 bp
57486:   contig of 6082 bp in length
63567:   gap of 100 bp
63668:   contig of 6429 bp in length
70096:   gap of 100 bp
70197:   contig of 7394 bp in length
77590:   gap of 100 bp
77691:   contig of 4569 bp in length
82259:   gap of 100 bp
82360:   contig of 7453 bp in length
89812:   gap of 100 bp
89913:   contig of 9198 bp in length
99110:   gap of 100 bp
99211:   contig of 9276 bp in length
108486:  gap of 100 bp
108587:  contig of 6439 bp in length
115025:  gap of 100 bp
115126:  contig of 25036 bp in length
140161:  gap of 100 bp
140262:  contig of 10416 bp in length
150677:  gap of 100 bp
150778:  contig of 12478 bp in length
163255:  gap of 100 bp
163356:  contig of 22743 bp in length
186098:  gap of 100 bp
186199:  contig of 10824 bp in length.
```

```
     misc_feature    15919..18437
                     /note="assembly_fragment"
     misc_feature    18538..20675
                     /note="assembly_fragment"
     misc_feature    20776..22054
                     /note="assembly_fragment"
     misc_feature    22155..25415
                     /note="assembly_fragment"
     misc_feature    25516..27880
                     /note="assembly_fragment"
     misc_feature    27981..30859
                     /note="assembly_fragment"
     misc_feature    30960..35486
                     /note="assembly_fragment"
     misc_feature    35587..38314
                     /note="assembly_fragment"
     misc_feature    38415..42074
                     /note="assembly_fragment"
     misc_feature    42175..45912
                     /note="assembly_fragment"
     misc_feature    46013..51604
                     /note="assembly_fragment"
     misc_feature    51705..57385
                     /note="assembly_fragment"
     misc_feature    57486..63567
                     /note="assembly_fragment"
     misc_feature    63668..70096
                     /note="assembly_fragment"
     misc_feature    70197..77590
                     /note="assembly_fragment"
     misc_feature    77691..82259
                     /note="assembly_fragment"
```

```
RESULT 13
AF429315/c
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
```

```
Query Match     5.4%;  Score 55.8;  DB 2;  Length 197022;
Best Local Similarity   70.1%;  Pred. No. 0.067;
Matches   89;  Conservative   0;  Mismatches   37;  Indels   1;  Gaps   1;

Qy     67 TCCCCATTTTTCAGATGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCAC 126
          ||||| |||||||||  |||| |  |  |||| |||||||||| || |||| || |||||||||
Db 160568 TCCCCATTTTTCAGGTGTGAAACTGAGGCCCAGAGAGAGGTTTAGTGACTTGCTGCAAGGTCAC 160627

Qy    127 ACAGCTAGAAAAGCAGCCAGGCC-CAGCCGAACCCCTGGGTGTGTGCAGCCCCCCAGCCCAGT 185
          ||||||  | |||||| |   || |  |  ||  ||| |||| | ||| || |||  ||||||
Db 160628 ACAGCTCAAAAGCAGCAAAAGCCCGAAGCTGAAACCCAGGTCTGTCCAGCTCCCTGGCCCAGG 160687

Qy    186 TGCTCAT 192
          ||| |||
Db 160688 CCATCTT 160694
```

FEATURES   Location/Qualifiers

Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

source   1..125020
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="16"
                /map="16q24.3; between D16S520 and WI-12410"
                /note="isolated from a patient with Huntington's
                Disease-Like 2 (HDL2)"

repeat_region   complement(35581..35746)
                /rpt_type=tandem
                /rpt_unit="ctg"

gene   complement(<36507..>36887)
                /gene="JPH3"
                /note="synonym: JP3"

mRNA   complement(<36507..36887)
                /gene="JPH3"
                /product="junctophilin 3"

CDS   complement(<36507..36887)
                /gene="JPH3"
                /note="component of the junctional complex between plasma
                membrane and endoplasmic reticulum"
                /codon_start=1
                /product="junctophilin 3"
                /protein_id="AAL40941.1"
                /db_xref="GI:17662245"
                /translation="MSSGRFNFDDGGSYCGWEDGKAHGHGVCTGPKGQGEYTGSWS
                HGFEVLGVVTWPSGNTYQGTWAQGKRHGIGLESKGKNVYKGEWTHGFGKGRYGVRECAG
                NGAKYEGTWSNGLQDGYGTETYSDG"

ORIGIN

Query Match   5.4%; Score 55.6; DB 9; Length 125020;
Best Local Similarity 13.5%; Pred. No. 0.084;
Matches 96; Conservative 301; Mismatches 308; Indels 8; Gaps 3;

Qy   214 GCGAGGCTGAGCAGCATGTGTTCCAGATGGTGGGAACTGGAGAGACCCCGGCCACAGGCCC 273
            :|| ::|:|:| :::||| | || :| ::|| | ::| |:: || :|||:::
Db   17564 GYYWGSKTCYSAGGKSRSMYYCMMRSSKSSSMSWSMSMARSSWCMGWGAGYRRSAGWGA 17505

Qy   274 GTGCAGGAGGAACCCGAGGGGCTGTAGGGCCCCGTGCCACTGCATGCCTCAGGGCTGGGTCC 333
            || |::||:| | :::| ::| ||| ::| |:| ||::||| :: | |||| ::
Db   17504 GWRSSKGRSTGMRACSKKTGSYGSTGRSMKKKGYSKYSRGMKGKKTCYCMWKYYKY 17445

Qy   334 TGGCAGCCACAGCCCCTTACTGCTGACCGCCAGCAGGAATCTGAGCCCCGGGAAGGGTCCAGG 393
            :|:|| :::|| || | ::|:||::|::| | |:| | |::|:|::|: |:| | ||
Db   17444 RKTSMCWWYYMKSWGYKYKRCCMKKKGCCTGYRGMSSKKSGYKSMRGMRGGSYSTSCWKS 17385

Qy   394 GAAGTTCGTGAACCATCTAGCAAGTCGGGCTGG--GGTGTGGCCCAAGTTGAGACACAGAATGT 452
            ::| :|| :: |||: :| |: | ::| |||| :||| |:::||||| |||:| | |||
Db   17444 CWGYSWMKCMKYMSYKKRRSMRGSMSSKGWRGYAGRGCYSSSMWSTRKRRSSKCYSYKSY 17325

Qy   453 AGGGCCCCTGGGA-------CTCAGAAATTGGCAGCTGCTCTTTTGGCCCCAGAGGGGCCCACCGCTG 507
            |:|| :::|| :|::| :| | : | : : | | | :|| :|:|::|::|| || |::
Db   17384 CWGYSWMKCMKYMSYKKRRSMRGSMSSKGWRGYAGRGCYSSSMWSTRKRRSSKCYSYKSY 17325

Qy   508 TGTCCGGGGCCTGGGTTAGCTCAGAAGGGTCACCTGGGGGGGTTCTTCCACTACACCCCCGCCTG 567
            |:||| :|::||:|:|::||| |::|: | ::| |||::|:|: | |||| |: |::|::||
Db   17324 KKGRGKMKGWGGMKRGSKYWSSMKKMKKRSSWSKCYSTKSYKSGRRKSKGWGRRSTKSKAKS 17265

Qy   568 GACACTGCTGTTAGCCCCAGGGCTCGAGGGACCACAGCTGGAGCGAGAGGAGGAGGAGGCCAG 627
            ::||| |:||:|:||:::||::::|:||::||:|::|:|||::|:|:|::|:|:|:|:| ||
Db   17264 SMRMAGSKCTYGSSYWSNNRNRNNRMGKTGCNYMYRRSRAMMNGNAAAAGCTTCCCCANTN 17205

Qy   628 TTCTCTCCTGTTAAGGGTATTTGCTTGTGAGCATGAGGGAACAGACAAGGCCACCCAGGGGACTAA 687
            :|:|:|::|:||:|:::|::||||| ||::|:|::||::||:|:|::|:|::|:::|::|| |
Db   17204 GGGGGAAAAGGCGGSGASRASCYKGRMSSKSCYRSGSTRRRCMKSKCRRGSRGKSMGMTRG 17145

Qy   688 CCCGAGATCCAGCCCCGGCCTCCGGCCCCTTTGGGGGGGGGTTGAGGGGGGTGAGGAGAAGCCCAGCAGCAGCAGTGG 807
            :|:|: :|||: :: |:|| | || ::|::::: ::|::| ::|:|:|:| |:|:|::|:| ||: |: | ::|
Db   17084 AMYCRSMCMCMCKSCCGCYSNGMSSYS--YSGKYSWGMKSYMWRSYSYSKRSTSKAWRSS 17027

Qy   748 TGAGCCTCCTGCCCAGCCTAGCAGGGGTTCCAGTGAGTGAGGGGGGTTGAGGGGGGTGAGGAGAAGCCCAGCAGCAGTGG 807

RESULT 14

LOCUS   AC021035   173791 bp   DNA   linear   PRI 20-APR-2002
DEFINITION   Homo sapiens chromosome 10 clone RP11-381K7, complete sequence.
ACCESSION   AC021035
VERSION   AC021035.8   GI:20219064
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 173791)
  AUTHORS   Smith,D.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Genome Therapeutics Corporation, 100 Beaver
             Street, Waltham, MA 02453, USA
REFERENCE   2 (bases 1 to 173791)
  AUTHORS   Smith,D.R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
             Sequence Data
             Genome Therapeutics Corporation Sequencing Center: Human Genome
REFERENCE   3 (bases 1 to 173791)
  AUTHORS   Smith,D.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
             Street, Waltham, MA 02453, USA
REFERENCE   4 (bases 1 to 173791)
  AUTHORS   Smith,D.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-APR-2002) Genome Therapeutics Corporation, 100 Beaver
             Street, Waltham, MA 02453, USA
COMMENT   On Apr 20, 2002 this sequence version replaced gi:18390167.
FEATURES   Location/Qualifiers
source   1..173791
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="RP11-381K7"
                /clone_lib="RPCI-11"

ORIGIN

Query Match   5.3%; Score 54.8; DB 9; Length 173791;
Best Local Similarity 55.9%; Pred. No. 0.11;
Matches 104; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy   45 TGTGGGTAGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGCCCAGAGCT 104
            |||||||||||||| | :| |||| |||||||::|||||:|:||||::|::|:||||
Db   98700 TGAGGTGGGGAGGGCTTTGTTCTTCCCATTTTTACAGAGAGAAAGTGAGGCTCAGGGAGT 98759

Qy   105 TAAGTAATCTGTCTGAGGGCCACACACCTAGAAAAGCAGCACAAGGCCAGGAACCCCTGGT 164
            ||||||| ||| |||||||::|:|:|:| ||||||:||||:|| :|::|::|::|::||
Db   98760 TAAGTAACCTGTCCAAGATCATGCAAATTAGAATGAGGCACACAACTTGAATTCAGGC 98819

Qy   165 GTGTGCAGGCCCCCAGCCCAGTTGCTCATTGCGGGGCTCGGGGAGCCCAGGCGAGGCTGAG 224
            |||||:|:|:|:|::||::|:||||| ||:||::|::|||:||:|:|::||::|:|:|
Db   98820 CTGGCTTACCCACAGAGCCTAGTGTTTGGGGTGTGTTGTTCTCACCCCAGGGCCTAGTCCAG 98879

Qy   225 CAGCAT 230
            ||||||
Db   98880 GGCCAT 98885

```
RESULT 15
ACO20911
LOCUS       ACO20911           196512 bp    DNA     linear    PRI 05-APR-2003
DEFINITION  Homo sapiens chromosome 19 clone CTD-2562J15, complete sequence.
ACCESSION   ACO20911
VERSION     ACO20911.9  GI:29568021
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. .
REFERENCE   1  (bases 1 to 196512)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196512)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 196512)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 196512)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 196512)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 5, 2003 this sequence version replaced gi:22128659.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.2% of Sequence;
            Estimated Total Number of Errors is 1.8.
            NOTE: 113bp single subclone 54713-54823.
FEATURES             Location/Qualifiers
     source          1..196512
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="CTD-2562J15"
                     /chromosome="19"
     misc_feature    54713..54823
                     /note="NOTE: 113bp single subclone 54713-54823."
ORIGIN

Query Match     5.1%;  Score 53.4; DB 9; Length 196512;
  Best Local Similarity   72.6%;
  Matches   69;  Conservative   0;  Mismatches   26;  Indels   0;  Gaps   0;

Qy         55 GGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGCCCCAGAGAACGTAAGTAATCT 114
              ||||  ||| ||| |||   ||| | |||| ||  |||| | |||| | |||| |||||
Db     139559 GGCTGTAGTTATTCCCATTTTACAGAGAGGGGTTGAGCCCCAGAGAGGGGCCAGGAAGTT 139618

Qy        115 GTCTGGAGGCCACACAGCTAGAAAGCAGCCAGGCCC 149
              || ||||| ||| || | ||| |||||||||| ||
Db     139619 GCCTGAAGCCACACACAGCAGCCCCAGAGGCCACGCCC 139653
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2005, 12:58:18 ; Search time 675 Seconds
                                     (without alignments)
                                     9103.247 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 cagggactgcgaagttcgcag.........ggcccctccccgccagcgcag 1038

Scoring table: IDENTITY NUC
                Gapop 10_0 , Gapext 1.0

Searched: 4390206 seqs, 295987O667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_16Dec04:*
                1: geneseqn1980s:*
                2: geneseqn1990s:*
                3: geneseqn200s:*
                4: geneseqn2001as:*
                5: geneseqn2001bs:*
                6: geneseqn2002as:*
                7: geneseqn2002bs:*
                8: geneseqn2003as:*
                9: geneseqn2003bs:*
                10: geneseqn2003cs:*
                11: geneseqn2003ds:*
                12: geneseqn2004as:*
                13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match	DB	ID	Description
1	1038	100.0	3006	12		Adj11355	Adj11355 Human PRS
c 2	55.8	5.4	111282	6		Aal44261	Aal44261 Human pho
c 3	55.8	5.4	111282	6		Abs55190	Abs55190 Genomic D
c 4	52.4	5.0	7086	5		Aal03024	Aal03024 Human rep
5	52.4	5.0	7086	5		Aal17415	Aal17415 Human rep
6	52.4	5.0	7388	5		Aba17415	Aba17415 Human rep
7	52.4	5.0	7388	5		Aal03025	Aal03025 Human ner
c 8	52.4	5.0	532	4		Aba17416	Aba17416 Human ner
c 9	49.6	4.8	2000	4		Aas32618	Aas32618 Human gen
c 10	49.2	4.8	414	4		Ada71938	Ada71938 Rice gene
c 11	49.2	4.7	7353	6		Aas32311	Aas32311 Human cDN
c 12	48.8	4.7	2000	8		Abn89472	Abn89472 Human G-p
c 13	48.2	4.6	29163	4		Ada71938	Ada71938 Rice gene
c 14	48.2	4.6	29163	4		Aal05121	Aal05121 Human rep
c 15	47.2	4.5	52661	10		Abl98013	Abl98013 Human tes
c 16	47.2	4.5	52661	10		Ada02876	Ada02876 Human DPT
c 17	47.2	4.5	52661	9		Adb72614	Adb72614 Human DPT
c 18	47.2	4.5	52661	12		Adc85355	Adc85355 Mouse Dpt
c 19	47.2	4.5	74822	6		Adm74471	Adm74471 Human car
c 20	47.2	4.5	99957	9		Abt10752	Abt10752 Human bre
c 21	47.2	4.5	99957	10		Adb72770	Adb72770 Human CA
c 22	47.2	4.5	99957	10		Adc85512	Adc85512 Human gen
c 23	47.2	4.5	99957	12		Adm74627	Adm74627 Human car
24	47	4.5	33923	4		Aak67071	Aak67071 Human imm
25	47	4.5	261817	11		Acn45182	Acn45182 Human gen
c 26	46.6	4.5	13216	4		Aal05122	Aal05122 Human rep
c 27	46.6	4.5	13216	4		Abl98014	Abl98014 Human tes
c 28	46.6	4.5	257645	12		Adq97289	Adq97289 Human can
c 29	46.2	4.5	703	4		Aac87363	Aac87363 Human dev
c 30	46.2	4.5	2849	10		Adb62282	Adb62282 Human cDN
c 31	46.2	4.4	63686	11		Acn44158	Acn44158 Human gen
c 32	45.4	4.4	243428	12		Adf51132	Adf51132 Human P-R
33	45	4.3	110000	12		Ado34435_3	Ado34435_3 Continuation (4 of
34	45	4.3	110000	13		Abd32806_5	Abd32806_5 Continuation (6 of
35	44.6	4.3	1741	4		Aah17445	Aah17445 Human cDN
36	44.6	4.3	1762	5		Aak87021	Aak87021 Human imm
37	44.6	4.3	1815	4		Aba19605	Aba19605 Human imm
38	44.6	4.3	1815	5		Aak87020	Aak87020 Human imm
39	44.6	4.3	57130	8		Aba19604	Aba19604 Human ner
40	44.6	4.3	57130	8		Abq77243	Abq77243 Human MAR
c 41	44.4	4.3	591	4		Aah09370	Aah09370 Human cDN
42	44.4	4.2	79597	10		Adl13866	Adl13866 Osteoarth
43	43.8	4.2	62118	11		Acn44566	Acn44566 Human gen
44	43.8	4.2	68001	12		Adj54477	Adj54477 Human B7-
45	43.8	4.2	223556	11		Acn44110	Acn44110 Human gen

ALIGNMENTS

RESULT 1
ADJ11355

ID   ADJ11355 standard; cDNA; 3006 BP.
XX
AC   ADJ11355;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Human PRSS11-L cDNA encoding a serine protease SeqID 1.
XX
KW   human; PRSS11-L; gene; ss; S2 serine protease; S2/HtrA;
KW   cellular physiology; apoptosis; gene therapy; osteopathic; stress;
KW   apoptotic; osteopathic; antiarthritic; tranquiliser.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   CDS   1011..2015
FT   /*tag= a
FT   /product= "PRSS11-L protein"
XX
PN   US2004005659-A1.
XX
PD   08-JAN-2004.
XX
PF   03-JUL-2002; 2002US-0189099.
XX
PR   03-JUL-2002; 2002US-0189099.
XX
PA   (DARR/) DARROW A L.
PA   (QlJ/) Qi J.
PA   (CHEN/) CHEN C.
PA   (ANDR/) ANDRADE-GORDON P.
XX
PI   Darrow AL, Qi J, Chen C, Andrade-Gordon P;
XX
PT   New isolated S2 serine protease nucleic acids and polypeptides, useful
PT   for diagnosing and/or treating diseases with aberrant expression or
PT   activity the S2 serine protease, such as osteoarthritis, stress and
PT   apoptotic disorders.

DR   WPI; 2004-081723/08.
DR   P-PSDB; ADJ11356.

```
XX
PS    Claim 2; SEQ ID NO 1; 28pp; English.
XX
CC    This invention relates to a novel isolated nucleic acid molecule
CC    identified as PRSS11-L that encodes an S2 serine protease. Specifically,
CC    it refers to members of the S2/HtrA serine protease family, such that it
CC    plays a role in cellular physiology and apoptosis. The present invention
CC    provides agonists, antagonists, antibodies and recombinant expression
CC    vectors useful in methods of treatment, or detection and diagnosis of
CC    diseases associated with the aberrant expression or activity of the S2
CC    serine protease, PRSS11-L. Accordingly, compositions described herein can
CC    be used via gene therapy routes to treat osteoarthritis, stress and
CC    apoptotic disorders. As such, they exhibit osteopathic, antiarthritic and
CC    tranquiliser activities. This polynucleotide sequence is the human PRSS11
CC    -L cDNA sequence of the invention.
XX
SQ    Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;

Query Match            100.0%;  Score 1038;  DB 12;  Length 3006;
Best Local Similarity  100.0%;
Matches 1038; Conservative 0;  Pred. No. 8.1e-260;
Matches 1038; Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy       1 CAGGGGACTCGAAGTTTGCGAGTCTGCAGTTCCCACACTCAGTTCCACAGATGGTGGTAGGGAGGAGGGCAT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAGGGGACTCGAAGTTTGCGAGTCTGCAGTTCCCACACTCAGTTCCACAGATGGTGGTAGGGAGGAGGGCAT 60

Qy      61 ATTCAGTGCCCATTTTTCAGATGAGGAGTTGAGCGGCCCAGAGAACGTAAGTAATCTGTCTGA 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATTCAGTGCCCATTTTTCAGATGAGGAGTTGAGCGGCCCAGAGAACGTAAGTAATCTGTCTGA 120

Qy     121 GGCCACACGCTAGAAAAGCAGCACAGGCCGCCCAGCCGCGAACCCCTGGTCTGTGTGCACGCCCCAGC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGCCACACGCTAGAAAAGCAGCACAGGCCGCCCAGCCGCGAACCCCTGGTCTGTGTGCACGCCCCAGC 180

Qy     181 CCAGTTGCTCATTGCGGGGCTCGGGAGCCACAGGCCACGAGGCACGAGCATGTGTTCCAGA 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCAGTTGCTCATTGCGGGGCTCGGGAGCCACAGGCCACGAGGCACGAGCATGTGTTCCAGA 240

Qy     241 TGGTGGGAACTGGAGAGAGCCCGCACCAGGCCCGTGCAGGGGAACCCCAGGGGCTGTAGGC 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGTGGGAACTGGAGAGAGCCCGCACCAGGCCCGTGCAGGGGAACCCCAGGGGCTGTAGGC 300

Qy     301 CCCGTGCCACTGCATGCCTCAGCGCTTGGTCTGGCAGCCACAGCGCCCCTACTGCTGACG 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCCGTGCCACTGCATGCCTCAGCGCTTGGTCTGGCAGCCACAGCGCCCCTACTGCTGACG 360

Qy     361 GCAGCAGGAATCTGAGCCCGGGAAGGGTTCCAGGGGAAGTTCGTGAACCATCTACCAAGTCG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCAGCAGGAATCTGAGCCCGGGAAGGGTTCCAGGGGAAGTTCGTGAACCATCTACCAAGTCG 420

Qy     421 GGCTGGGGGTGTGGCCCAAGTTAGACACAGATGTAGGGGCCCCTGGGACTCAGAAAATTGGCCAG 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGCTGGGGGTGTGGCCCAAGTTAGACACAGATGTAGGGGCCCCTGGGACTCAGAAAATTGGCCAG 480

Qy     481 CTCTTTTTGGCCCAGAGGGGCCCACGCGTTGTCTCGGGGTAGCTCAGAAGGGTCACCT 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTCTTTTTGGCCCAGAGGGGCCCACGCGTTGTCTCGGGGTAGCTCAGAAGGGTCACCT 540

Qy     541 GGGGGTCTTCCACTACACCCCGCCCCGCCCCTGGACACTGCCTCGTAGCCTCGGAGGGGACC 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGGGGTCTTCCACTACACCCCGCCCCGCCCCTGGACACTGCCTCGTAGCCTCGGAGGGGACC 600

Qy     601 AGCTGGAGCCCATGAGGAGGAGGGCCAGTTCTCTCCTGTAAGGGTATTGCTGTGTAGCCATGAG 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGCTGGAGCCCATGAGGAGGAGGGCCAGTTCTCTCCTGTAAGGGTATTGCTGTGTAGCCATGAG 660

Qy     661 GGAAACAGACAAGGCCCAGGGGGACTAACCCGAGATCCAGCCCCGCGAGGCCTCACTCCCGTGTG 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGAAACAGACAAGGCCCAGGGGGACTAACCCGAGATCCAGCCCCGCGAGGCCTCACTCCCGTGTG 720

Qy     721 GCTCACGCGGCAAATATCCTAACCTCTCTCTGAGCCTTCGAGGCCTTAGCAGGGGTCCAGTG 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCTCACGCGGCAAATATCCTAACCTCTCTCTGAGCCTTCGAGGCCTTAGCAGGGGTCCAGTG 780

Qy     781 AGGGGGGTTGAGGAAGCCAGCACGTGGGAAGCCTTTTTTAACCATTCTCGGGGTGAGCGAGC 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGGGGGGTTGAGGAAGCCAGCACGTGGGAAGCCTTTTTTAACCATTCTCGGGGTGAGCGAGC 840

Qy     841 CCCTTCCCAAATGCCTCGTGTCACTGCACTGCTTGTGTGGTAGGGGGTCCCCAGGGGGCTC 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCCTTCCCAAATGCCTCGTGTCACTGCACTGCTTGTGTGGTAGGGGGTCCCCAGGGGGCTC 900

Qy     901 AGTGTGGGGCTGAGGCTGCGTGCGGACGCAGGGGGTGTCTCAGGAAGAGAGCCCTCTCCC 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGTGTGGGGCTGAGGCTGCGTGCGGACGCAGGGGGTGTCTCAGGAAGAGAGCCCTCTCCC 960

Qy     961 TGCCCACTGGGGCATAGGGCCTCTGGGAGAGCTGGAGGAGCATCGTGATCTGATCTGATCTCACTGATCGCACCTGG 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TGCCCACTGGGGCATAGGGCCTCTGGGAGAGCTGGAGGAGCATCGTGATCTGATCTGATCTCACTGATCGCACCTGG 1020

Qy    1021 CCCTTCCCGCCCACGCCAG 1038
           |||||||||||||||||||
Db    1021 CCCTTCCCGCCCACGCCAG 1038


RESULT 2
AAL44261/c
ID   AAL44261   standard; DNA; 111282 BP.
XX
AC   AAL44261;
XX
DT   24-OCT-2002  (first entry)
XX
DE   Human phosphodiesterase protein (PDE) gene sequence.
XX
KW   Human; gene; ds; gene therapy; vaccine; protein therapy;
KW   phosphodiesterase protein; PDE; PDE2A; asthma;
KW   chronic obstructive pulmonary disease; rheumatoid arthritis; atopy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(254, A)
FT                   /*tag= a
FT   variation       replace(1997, T)
FT                   /standard_name=
FT                   /*tag= b
FT   variation       replace(2653, A)
FT                   /standard_name=
FT                   /*tag= c
FT   variation       replace(2986, A)
FT                   /standard_name=
FT                   /*tag= d
FT   variation       replace(3702, A)
FT                   /standard_name=
FT                   /*tag= e
FT   variation       replace(3710, G)
FT                   /standard_name=
FT                   /*tag= f
FT   variation       replace(4461, T)
FT                   /standard_name=
FT                   /*tag= g
FT   variation       replace(4734, A)
FT                   /standard_name=
FT                   /*tag= h
FT   variation       replace(5649, G)
FT                   /standard_name=
FT                   /*tag= i
FT   variation       replace(5957, G)
FT                   /standard_name=
FT                   /*tag= j
FT   variation       replace(6060, C)
FT                   /standard_name=
FT                   /*tag= k
FT   variation       replace(6133, G)
FT                   /standard_name=
```

Qy 781 "Single nucleotide polymorphism"
Qy 841 "Single nucleotide polymorphism"
Qy 901 "Single nucleotide polymorphism"
Qy 961 "Single nucleotide polymorphism"

```
FT   variation
FT                   /replace(6692, G/C/T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= l
FT   variation
FT                   /replace(6911, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= m
FT   variation
FT                   /replace(7541, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= n
FT   variation
FT                   /replace(9072, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= o
FT   variation
FT                   /replace(8173, T/C/A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= p
FT   variation
FT                   /replace(8694, G/T/C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= q
FT   variation
FT                   /replace(9426, G/C/T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= r
FT   variation
FT                   /replace(9573, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= s
FT   variation
FT                   /replace(9826, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= t
FT   variation
FT                   /replace(10134, T/A/G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= u
FT   variation
FT                   /replace(11014, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= v
FT   variation
FT                   /replace(12390, A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= w
FT   variation
FT                   /replace(14701, T/A/G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= x
FT   variation
FT                   /replace(15679, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= y
FT   variation
FT                   /replace(18606, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= z
FT   variation
FT                   /replace(19070, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= aa
FT   variation
FT                   /replace(19470, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ab
FT   variation
FT                   /replace(19611, A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ac
FT   variation
FT                   /replace(20641, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ad
FT   variation
FT                   /replace(20642, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ae
FT   variation
FT                   /replace(21035, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= af
FT   variation
FT                   /replace(21871, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ag
FT   variation
FT                   /replace(22907, A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ah
FT   variation
FT                   /replace(23722, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ai
FT   variation
FT                   /replace(23722, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= aj

FT   variation
FT                   /replace(24121, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ak
FT   variation
FT                   /replace(24553, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= al
FT   variation
FT                   /replace(25917, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= am
FT   variation
FT                   /replace(27525, A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= an
FT   variation
FT                   /replace(27625, C/A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ao
FT   variation
FT                   /replace(27833, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ap
FT   variation
FT                   /replace(27852, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= aq
FT   variation
FT                   /replace(28478, A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ar
FT   variation
FT                   /replace(28514, C/G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= as
FT   variation
FT                   /replace(28702, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= at
FT   variation
FT                   /replace(28859, C/G/T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= au
FT   variation
FT                   /replace(28960, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= av
FT   variation
FT                   /replace(29030, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= aw
FT   variation
FT                   /replace(29348, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ax
FT   variation
FT                   /replace(29973, C/G/A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ay
FT   variation
FT                   /replace(30153, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= az
FT   variation
FT                   /replace(30389, C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= ba
FT   variation
FT                   /replace(30581, G)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= bb
FT   variation
FT                   /replace(31147, A)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= bc
FT   variation
FT                   /replace(31224, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= bd
FT   variation
FT                   /replace(31735, A/T/C)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= be
FT   variation
FT                   /replace(31798, C/A/T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= bf
FT   variation
FT                   /replace(31994, G/T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= bg
FT   variation
FT                   /replace(32324, T)
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /*tag= bh
```

```
FT   variation             replace(32891, A)
FT                         /*tag= b1
FT   variation             /standard_name= "Single nucleotide polymorphism"
FT                         replace(33104, T)
FT                         /*tag= bj

Qy   Query Match           5.4%; Score 55.8; DB 6; Length 111282;
     Best Local Similarity 70.1%; Pred. No. 0.001;
     Matches 89; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

Qy      67 TCCCATTTTTCAGATGAGGAGGTTGAGGCCCAGAGAAGTAAGTAAACTGTTCTGAGGCCAC 126
           |||||||||||||| ||||||||||||||| | | | ||||||| ||||||||||||||
Db   93162 TCCCATTTTTCAGGTGTGAAAACTGAGGCCCCAGAGAGGTTTAGTGACTTGCTCAAAGGTCAC 93103

Qy     127 ACAGCTAGAAAAACAGCAGGGCC-CAGCCGAACCCTGGTGTGTGTGCAGCCCCAGCCCAGT 185
           ||||||||||||||||||||||   |||  |||||||||||||||||||||||||||||||
Db   93102 ACAGCTCAAAAACAGCAAAACCGAAGTGTCTGTCCAGCTTCCCTGGCCAGG 93043

Qy     186 TGCTTCAT 192
           ||||||||
Db   93042 CCATCTT 93036


RESULT 3
ABSS5190/c
ID   ABSS5190 standard; DNA; 111282 BP.
XX
AC   ABSS5190;
XX
DT   16-DEC-2002 (first entry)
XX
DE   Genomic DNA encoding human phosphodiesterase (PDE) protein.
XX
KW   Human; phosphodiesterase; PDE; phosphodiesterase 2A; PDE2A; gene therapy;
KW   chromosome 11; gene; ds; single nucleotide polymorphism; SNP;
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   variation             replace(254, A)
FT                         /*tag= a
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(1997, T)
FT                         /*tag= b
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(2653, A)
FT                         /*tag= c
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(2986, A)
FT                         /*tag= d
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(3702, A)
FT                         /*tag= e
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(3710, A)
FT                         /*tag= f
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(4461, T)
FT                         /*tag= g
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(4734, A)
FT                         /*tag= h
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(5549, G)
FT                         /*tag= i
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(5967, C)
FT                         /*tag= j
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(6060, C)
FT                         /*tag= k
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(6613, G)
FT                         /*tag= l
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"

FT                         /*tag= 1
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(6692, H)
FT                         /*tag= m
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(6911, G)
FT                         /*tag= n
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(7541, T)
FT                         /*tag= o
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(8173, H)
FT                         /*tag= p
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(8694, B)
FT                         /*tag= q
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(9072, G)
FT                         /*tag= r
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(9425..9426, A)
FT                         /*tag= s
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(9425..9426, AA)
FT                         /*tag= t
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(9425..9426, AAA)
FT                         /*tag= u
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(9426, H)
FT                         /*tag= v
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(9826, C)
FT                         /*tag= w
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(10134, V)
FT                         /*tag= x
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(11014, T)
FT                         /*tag= y
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(12390, A)
FT                         /*tag= z
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(13719..13721, TT)
FT                         /*tag= aa
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(14701, V)
FT                         /*tag= ab
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(15678..15680, GT)
FT                         /*tag= ac
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(15686..15688, TG)
FT                         /*tag= ad
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(18321..18323, CT)
FT                         /*tag= ae
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(18606, C)
FT                         /*tag= af
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(19470, G)
FT                         /*tag= ag
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(19611, A)
FT                         /*tag= ah
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(19707, G)
FT                         /*tag= ai
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation             replace(20641, C)
FT                         /*tag= aj
FT                         /standard_name= "Single nucleotide polymorphism (SNP)"
```

```
FT   variation
FT            /replace(20642,A)
FT            /*tag= ak
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(21036,T)
FT            /*tag= al
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(21871,C)
FT            /*tag= am
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(22907,A)
FT            /*tag= an
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(23722,G)
FT            /*tag= ao
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(24121,C)
FT            /*tag= ap
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(24553,C)
FT            /*tag= aq
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(25917,C)
FT            /*tag= ar
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(26572..26574,CG)
FT            /*tag= as
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(27525,A)
FT            /*tag= at
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(27625,Y)
FT            /*tag= au
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(27833,C)
FT            /*tag= av
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(27852,G)
FT            /*tag= aw
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(28478,C)
FT            /*tag= ax
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(28514,Y)
FT            /*tag= ay
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(28702,C)
FT            /*tag= az
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(28859,H)
FT            /*tag= ba
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(28960,G)
FT            /*tag= bb
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(29030,G)
FT            /*tag= bc
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(29348,C)
FT            /*tag= bd
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(29973,H)
FT            /*tag= be
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(30153,T)
FT            /*tag= bf
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(30389,C)
FT            /*tag= bg
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(30581,C)
FT            /*tag= bh
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(31147,G)
FT            /*tag= bi
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation
FT            /replace(31224,C)
FT            /*tag= bj
FT            /standard_name= "Single nucleotide polymorphism (SNP)"
XX
XX
Qy          67 TCCCATTTTTCAGATGAGGAGTTGAGCCCAGAGAACGTAAGTAAATCTGTCTGAGGCCAC 126
               |||||||||||||||| ||||||||| ||||||||| ||| ||||||||||| || |||||
Db       93162 TCCCCATTTTCAGGTGTGAAAACTGAGCGCCCAGAGAGGTTTAGTTGACTTGCTCAAGGTCAC 93103
Qy         127 ACAGCTAGAAAAGCAGCAGCCGGCC-CAGCCGGAACCCCTGGTGTGGTTGCAGCCCCCAGCCCAGT 185
               |||| |||||||| ||||   |||| || |   |||||| ||| |||| |||| |||||| |
Db       93102 ACAGCCTCAAAAACCAGCAAAAGCCGAAGCTGAAACCCAGGTCTGTCCAGCTCCCTGGCCAGG 93043
Qy         186 TGCTCAT 192
               |||||||
Db       93042 CCATCTT 93036
```

RESULT 4
AAL03024
ID   AAL03024 standard; DNA; 7086 BP.
XX
AC   AAL03024;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 5712.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189074P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   14-AUG-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0244518P.
PR   14-AUG-2000; 2000US-0245199P.
PR   14-AUG-2000; 2000US-0245213P.
PR   14-AUG-2000; 2000US-0252514P.
PR   14-AUG-2000; 2000US-0252267P.
PR   14-AUG-2000; 2000US-0252267P.
PR   14-AUG-2000; 2000US-0252268P.
PR   14-AUG-2000; 2000US-0252269P.
PR   14-AUG-2000; 2000US-0252270P.
PR   14-AUG-2000; 2000US-0255447P.
PR   14-AUG-2000; 2000US-0255757P.

Query Match   5.4%; Score 55.8; DB 6; Length 111282;
Best Local Similarity   70.1%; Pred. No. 0.001;
Matches   89; Conservative   0; Mismatches   37; Indels   1; Gaps   1;

```
PR   14-AUG-2000;  2000US-0225758P.
PR   14-AUG-2000;  2000US-0225759P.
PR   18-AUG-2000;  2000US-0226279P.
PR   22-AUG-2000;  2000US-0226668P.
PR   22-AUG-2000;  2000US-0226868P.
PR   23-AUG-2000;  2000US-0227182P.
PR   22-AUG-2000;  2000US-0227008P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   05-SEP-2000;  2000US-0230437P.
PR   06-SEP-2000;  2000US-0230438P.
PR   06-SEP-2000;  2000US-0231242P.
PR   08-SEP-2000;  2000US-0231243P.
PR   08-SEP-2000;  2000US-0231244P.
PR   08-SEP-2000;  2000US-0231413P.
PR   08-SEP-2000;  2000US-0231414P.
PR   08-SEP-2000;  2000US-0232080P.
PR   08-SEP-2000;  2000US-0232081P.
PR   12-SEP-2000;  2000US-0231968P.
PR   14-SEP-2000;  2000US-0232397P.
PR   14-SEP-2000;  2000US-0232398P.
PR   14-SEP-2000;  2000US-0232399P.
PR   14-SEP-2000;  2000US-0232400P.
PR   14-SEP-2000;  2000US-0232401P.
PR   14-SEP-2000;  2000US-0233063P.
PR   14-SEP-2000;  2000US-0233064P.
PR   14-SEP-2000;  2000US-0233065P.
PR   21-SEP-2000;  2000US-0234233P.
PR   21-SEP-2000;  2000US-0234234P.
PR   25-SEP-2000;  2000US-0234997P.
PR   25-SEP-2000;  2000US-0234998P.
PR   26-SEP-2000;  2000US-0234984P.
PR   27-SEP-2000;  2000US-0235834P.
PR   27-SEP-2000;  2000US-0235836P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236367P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   02-OCT-2000;  2000US-0239935P.
PR   13-OCT-2000;  2000US-0239377P.
PR   13-OCT-2000;  2000US-0239937P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241221P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241786P.
PR   20-OCT-2000;  2000US-0241787P.
PR   20-OCT-2000;  2000US-0241808P.
PR   20-OCT-2000;  2000US-0241809P.
PR   20-OCT-2000;  2000US-0241826P.
PR   01-NOV-2000;  2000US-0244617P.
PR   01-NOV-2000;  2000US-0246474P.
PR   08-NOV-2000;  2000US-0464766P.
PR   08-NOV-2000;  2000US-0464767P.
PR   08-NOV-2000;  2000US-0464777P.
PR   08-NOV-2000;  2000US-0464778P.
PR   08-NOV-2000;  2000US-0465523P.
PR   08-NOV-2000;  2000US-0465524P.
PR   08-NOV-2000;  2000US-0465525P.
PR   08-NOV-2000;  2000US-0465526P.
PR   08-NOV-2000;  2000US-0465527P.
PR   08-NOV-2000;  2000US-0465528P.
PR   08-NOV-2000;  2000US-0465532P.
PR   08-NOV-2000;  2000US-0466609P.
PR   08-NOV-2000;  2000US-0246610P.
PR   08-NOV-2000;  2000US-0246611P.
PR   08-NOV-2000;  2000US-0246613P.
PR   17-NOV-2000;  2000US-0249207P.
PR   17-NOV-2000;  2000US-0249208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251031P.
PR   05-DEC-2000;  2000US-0251988P.
PR   06-DEC-2000;  2000US-0251479P.
PR   06-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251857P.
PR   08-DEC-2000;  2000US-0251858P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 5712; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 7086 BP; 1264 A; 2050 C; 1979 G; 1793 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 52.4;  DB 4;  Length 7086;
  Best Local Similarity   69.6%;
  Matches   71;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy         68 CCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAATCTGTTCTGAGGCCACA 127
              | ||||||  ||||| |  | ||||||| |  || | ||||||||||||   || |||||||
Db       4323 CTCATTTTATAGATGAAACACTCGGGGCCTTAGAGAAGATAAGTAATCTCCCAAGGCCACA 4382

Qy        128 CAGCTAGAAAGCAGCCAGCCGACCCTGGGTGTGTG 169
              |||||||| ||| ||| |  | |||  || |||||
Db       4383 CAGCTAGGAAGTAGCAGAGCTTATGCCACGTCAGGGGTTTGTG 4424

RESULT 5
ABA17415
ID   ABA17415 standard; DNA; 7086 BP.
XX
AC   ABA17415;
```

XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9746.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO20015063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX

PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0191123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220964P.
PR   26-JUL-2000; 2000US-0220963P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0271182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229343P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   08-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0320080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.

PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236277P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   02-OCT-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   13-OCT-2000; 2000US-0237040P.
PR   20-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0242221P.
PR   20-OCT-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0244474P.
PR   08-NOV-2000; 2000US-0244475P.
PR   08-NOV-2000; 2000US-0244476P.
PR   08-NOV-2000; 2000US-0244477P.
PR   08-NOV-2000; 2000US-0244478P.
PR   08-NOV-2000; 2000US-0245523P.
PR   08-NOV-2000; 2000US-0245524P.
PR   08-NOV-2000; 2000US-0245525P.
PR   08-NOV-2000; 2000US-0245526P.
PR   08-NOV-2000; 2000US-0245527P.
PR   08-NOV-2000; 2000US-0245528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250391P.
PR   01-DEC-2000; 2000US-0251160P.

```
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
PT   and metastases.
XX
PS   Disclosure; SEQ ID NO 9746; 170lpp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note:The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7086 BP; 1264 A; 2050 C; 1979 G; 1793 T; 0 U; 0 Other;

Query Match             5.0%;  Score 52.4;  DB 5;  Length 7086;
Best Local Similarity   69.6%;
Matches    71;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          68 CCCATTTTTCAGATGAGAGAGTTGAGGCCCAGAGAACGTAAGTAAATCTGTCTGAGGCCACA 127
               |||||||  |||| ||||| ||| |||| ||||| || ||||||||||||| ||||||||||
Db        4323 CTCATTTTATAGATGAAACACTGGGCCCTAGAGAGAGATAAGTAAATCATCTCCAAGGCCACA 4382

Qy         128 CAGCTAGAAAGCAGCAGCCCAGCCCGAACCCCTGGGTCGTGTG 169
               ||||||||||||||| |||||||||||||||||||||||||||
Db        4383 CAGCTAGGAAGTAGCAGAGCTTATGCCACGTCAGGGTTTGTG 4424
```

```
RESULT 6
AAL03025
ID   AAL03025 standard; DNA; 7388 BP.
XX
AC   AAL03025;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 5713.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
```

```
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0245188P.
PR   14-AUG-2000; 2000US-0245198P.
PR   14-AUG-2000; 2000US-0245213P.
PR   14-AUG-2000; 2000US-0245214P.
PR   14-AUG-2000; 2000US-0252144P.
PR   14-AUG-2000; 2000US-0252145P.
PR   14-AUG-2000; 2000US-0252467P.
PR   14-AUG-2000; 2000US-0252617P.
PR   14-AUG-2000; 2000US-0252668P.
PR   14-AUG-2000; 2000US-0252688P.
PR   14-AUG-2000; 2000US-0252689P.
PR   14-AUG-2000; 2000US-0252700P.
PR   14-AUG-2000; 2000US-0254417P.
PR   14-AUG-2000; 2000US-0254477P.
PR   14-AUG-2000; 2000US-0255757P.
PR   14-AUG-2000; 2000US-0255758P.
PR   14-AUG-2000; 2000US-0255759P.
PR   18-AUG-2000; 2000US-0262079P.
PR   22-AUG-2000; 2000US-0262799P.
PR   22-AUG-2000; 2000US-0266681P.
PR   22-AUG-2000; 2000US-0266868P.
PR   22-AUG-2000; 2000US-0271782P.
PR   23-AUG-2000; 2000US-0277009P.
PR   30-AUG-2000; 2000US-0288924P.
PR   01-SEP-2000; 2000US-0292877P.
PR   01-SEP-2000; 2000US-0293443P.
PR   01-SEP-2000; 2000US-0293444P.
PR   01-SEP-2000; 2000US-0293455P.
PR   05-SEP-2000; 2000US-0295099P.
PR   05-SEP-2000; 2000US-0295512P.
PR   06-SEP-2000; 2000US-0304377P.
PR   06-SEP-2000; 2000US-0304388P.
PR   08-SEP-2000; 2000US-0312429P.
PR   08-SEP-2000; 2000US-0312439P.
PR   08-SEP-2000; 2000US-0312444P.
PR   08-SEP-2000; 2000US-0314413P.
PR   08-SEP-2000; 2000US-0314414P.
PR   08-SEP-2000; 2000US-0320080P.
PR   08-SEP-2000; 2000US-0320081P.
PR   12-SEP-2000; 2000US-0321968P.
PR   14-SEP-2000; 2000US-0323397P.
PR   14-SEP-2000; 2000US-0323398P.
PR   14-SEP-2000; 2000US-0323399P.
PR   14-SEP-2000; 2000US-0324001P.
PR   14-SEP-2000; 2000US-0324401P.
PR   14-SEP-2000; 2000US-0330637P.
PR   14-SEP-2000; 2000US-0330639P.
PR   14-SEP-2000; 2000US-0330658P.
PR   21-SEP-2000; 2000US-0342223P.
PR   21-SEP-2000; 2000US-0342274P.
PR   25-SEP-2000; 2000US-0349907P.
PR   25-SEP-2000; 2000US-0349988P.
PR   26-SEP-2000; 2000US-0354449P.
PR   27-SEP-2000; 2000US-0354954P.
PR   27-SEP-2000; 2000US-0355834P.
PR   27-SEP-2000; 2000US-0358836F.
PR   29-SEP-2000; 2000US-0336327P.
```

DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 5713; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention
XX
SQ   Sequence 7388 BP; 1347 A; 2127 C; 2078 G; 1836 T; 0 U; 0 Other;

     Query Match   5.0%;   Score 52.4;   DB 4;   Length 7388;
     Best Local Similarity   69.6%;   Pred. No. 0.0036;
     Matches   71;   Conservative   0;   Mismatches   31;   Indels   0;   Gaps   0;

Qy   68 CCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACCTAAGTAAATCTGTTCTGAGGCCACA 127
            |||||| |||| ||||| |||| |||||| |||||| ||||||| || ||||||||||
Db   4327 CTCATTTTATAGATGAAACACTGGGGCCTAGAGAAGATAAGTAAATCTCTCCAAGGCCACA 4386

Qy   128 CAGCTAGAAAAGCAGCCAGGCCCAGCCGAACCCCTGGTGTGTG 169
            ||||||| ||| ||||| ||| || |||| ||| ||||||
Db   4387 CAGCTAGGAAGTAGCAGAGCTTATGCCACGTCAGGGTTTGTG 4428

RESULT 7
ABA17416
ID   ABA17416 standard; DNA; 7388 BP.
XX
AC   ABA17416;
XX
DT   23-JAN-2002   (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9747.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide; immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary; antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal; antiparasitic; cardiant; immune disorder; cardiovascular disorder; neurological disease; infection; nephrotropic; gene therapy; vaccine; de.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216477P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217496P.
PR   11-JUL-2000; 2000US-0217497P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.

PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
XX
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
XX
PR   02-OCT-2000; 2000US-0370040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0418026P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0244674P.
PR   08-NOV-2000; 2000US-0244675P.
PR   08-NOV-2000; 2000US-0244676P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   08-NOV-2000; 2000US-0492007P.
PR   17-NOV-2000; 2000US-0492008P.
PR   17-NOV-2000; 2000US-0492009P.
PR   17-NOV-2000; 2000US-0492010P.
PR   17-NOV-2000; 2000US-0492011P.
PR   17-NOV-2000; 2000US-0492012P.
PR   17-NOV-2000; 2000US-0492013P.
PR   17-NOV-2000; 2000US-0492214P.
PR   17-NOV-2000; 2000US-0492215P.
PR   17-NOV-2000; 2000US-0492216P.
PR   17-NOV-2000; 2000US-0492217P.
PR   17-NOV-2000; 2000US-0492218P.
PR   17-NOV-2000; 2000US-0492244P.
PR   17-NOV-2000; 2000US-0492245P.
PR   17-NOV-2000; 2000US-0492264P.
PR   17-NOV-2000; 2000US-0492265P.
PR   17-NOV-2000; 2000US-0492297P.
PR   17-NOV-2000; 2000US-0492299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   24-FEB-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251479P.
PR   06-DEC-2000; 2000US-0256719P.
PR   08-DEC-2000; 2000US-0251990P.
PR   08-DEC-2000; 2000US-0251989P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0256078P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0271182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 08-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234222P.
PR 14-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237027P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0242617P.
PR 01-NOV-2000; 2000US-0244611P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0465527P.
PR 08-NOV-2000; 2000US-0465528P.
PR 08-NOV-2000; 2000US-0465532P.
PR 08-NOV-2000; 2000US-0466609P.
PR 08-NOV-2000; 2000US-0466610P.
PR 08-NOV-2000; 2000US-0466611P.
PR 08-NOV-2000; 2000US-0466613P.
PR 17-NOV-2000; 2000US-0492072P.
PR 17-NOV-2000; 2000US-0492092P.
PR 17-NOV-2000; 2000US-0492097P.
PR 17-NOV-2000; 2000US-0492102P.
PR 17-NOV-2000; 2000US-0492111P.
PR 17-NOV-2000; 2000US-0492112P.
PR 17-NOV-2000; 2000US-0492113P.
PR 17-NOV-2000; 2000US-0492114P.
PR 17-NOV-2000; 2000US-0492115P.
PR 17-NOV-2000; 2000US-0492116P.
PR 17-NOV-2000; 2000US-0492117P.
PR 17-NOV-2000; 2000US-0492180P.
PR 17-NOV-2000; 2000US-0492244P.
PR 17-NOV-2000; 2000US-0492245P.
PR 17-NOV-2000; 2000US-0492264P.
PR 17-NOV-2000; 2000US-0492265P.
PR 17-NOV-2000; 2000US-0492297P.
PR 17-NOV-2000; 2000US-0492299P.
PR 17-NOV-2000; 2000US-0492997P.
PR 01-DEC-2000; 2000US-0503091P.
PR 01-DEC-2000; 2000US-0511600P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251479P.
PR 05-DEC-2000; 2000US-0267199P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 17-NOV-2000; 2000US-0249300P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
WPI; 2001-541565/60.
DR
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases.
PT
XX
PS Disclosure; SEQ ID NO 9747; 170lpp + Sequence Listing; English.
XX

CC The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABB14678-ABB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly

```
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7388 BP; 1347 A; 2127 C; 2078 G; 1836 T; 0 U; 0 Other;

Query Match    5.0%;  Score 52.4;  DB 5;  Length 7388;
Best Local Similarity   69.6%;  Pred. No. 0.0036;
Matches   71;  Conservative   0;  Mismatches   31;  Indels   0;  Gaps   0;

Qy      68 CCCATTTTCAGATGAGGAGGTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCACA 127
           |||||||   | ||||    ||||||| ||  || || || ||||||  ||||||||||||
Db    4327 CTCATTTTATAGATGAAACACTGGGGGCCTAGAGAAGATAAGTAAATCTCTCCAAGGCCACA 4386

Qy     128 CAGCTAGAAAGCAGCCAGGCCCAGCCGGAACCCCTGTCTGTG 169
           |||||||| || |||| |||      |      || |||||||
Db    4387 CAGCTAGAAGTAGCAGAGAGCTTATGCCACGTCAGGGGTTGTG 4428
```

```
RESULT 8
AAS32618/c
ID   AAS32618 standard; DNA; 532 BP.
XX
AC   AAS32618;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human genomic DNA for novel endocrine antigen, SEQ ID No 572.
XX
KW   Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW   thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW   antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW   reproductive disorder; endocrine cancer; pancreatic disorder;
KW   diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW   hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200155319-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001335.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   11-JUL-2000; 2000US-0217496P.
PR   11-JUL-2000; 2000US-0217487P.
PR   07-JUL-2000; 2000US-0216880P.
PR   07-JUL-2000; 2000US-0216647P.
PR   30-JUN-2000; 2000US-0215135P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0220963P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
```

```
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227099P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230477P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   26-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   27-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239377P.
PR   20-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   01-NOV-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   20-OCT-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
```

```
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   01-DEC-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   05-DEC-2000; 2000US-0249245P.
PR   05-DEC-2000; 2000US-0249265P.
PR   06-DEC-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-457726/49.
XX
PT   Isolated polypeptide for treating, preventing and prognosing disorders
PT   related to the endocrine, and gastrointestinal disorders and also for
PT   testing and detection e.g. diagnosis.
XX
PS   Disclosure; SEQ ID NO 572; 558pp; English.
XX
CC   The invention relates to cDNAs encoding novel human endocrine antigens or
CC   a fragment having biological activity, a domain, an epitope, full length
CC   protein, variant, allelic variant or a species homologue of the
CC   cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC   treating or ameliorating a medical condition when administered (e.g. by
CC   gene therapy or antisense-therapy). Identifying mutations in the genes
CC   coding for the antigens is useful for diagnosing a pathological condition
CC   or a susceptibility to a pathological condition. The DNAs, antigens and
CC   antibodies raised against the antigens useful for treating, preventing
CC   and/or prognosing disorders related to the endocrine system or hormone
CC   imbalance or reproductive disorders, cancers of endocrine tissues,
CC   disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC   (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC   hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC   of diseases and disorders are given in the specification. The present
CC   sequence is genomic DNA fragment form a gene encoding an endocrine
CC   antigen of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 532 BP; 124 A; 164 C; 121 G; 123 T; 0 U; 0 Other;

Query Match             4.8%;  Score 49.6;  DB 4;  Length 532;
Best Local Similarity   65.2%;  Pred. No. 0.009;
Matches   73;  Conservative    0;  Mismatches  39;  Indels    0;  Gaps    0;

Qy          67 TCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAAACGTAAGTAATTCGTCTGAGGCCAC 126
```

```
Db         197 TGCCCTTTTAGAGATGAAGGAAGGACTGAGGCTCAGAGAAGGCCAAGTAATTCACCTAAGGCCAC 138
                |||| |||||||| ||  ||| ||||| ||| |||||  | ||||||||  ||||||||||||
Qy         127 ACAGCTAGAAAAGCCAGCCCAGGCGCACCCCTGGTGTGTGTGCAGCCGTT 178
                ||||    ||||||||||  ||  |||||  |  |||||||||||||
Db         137 GTAGCTAGCAGAGAAAGCCGAGACTCAACCTAAGTCTGAGGCTCCTGCCA 86
```

```
RESULT 9
ADA71938
ID   ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
DR   WPI; 2003-175290/17.
XX
PN   WO2003000898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H,  Chen W,  Cooper B,  Glazebrook J,  Goff SA,  Hou Y;
PI   Katagiri F,  Quan S,  Tao Y,  Whitham S,  Xie Z,  Zhu T,  Zou G;
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. The gene is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match             4.8%;  Score 49.4;  DB 8;  Length 2000;
Best Local Similarity   61%;  Pred. No. 0.015;
Matches   61;  Conservative  294;  Mismatches  295;  Indels    1;  Gaps    1;

Qy         114 TGTCTGAGGCCACACAGCTAGAAAAGCCAGCCCAGGCGCACCCCTGGTGTGTGTGCAGC 173
                | ||| ||| |:  |:||:| :|: |: |:|: |: || :|| ||  |:|:||:|||
Db          94 YKRSRWRGRGRRGMRRSRMRWRGRYRRCARSGRMAGGSGRMWGGKSRMSYWMWCYARGC 153
```

```
Qy         174 CCCCAGCCCAGTTGCTTCATTGCGGGGGGCTCGGGAGGCCACGAGCCGAGGCTGAGCAGCATGTG 233
                |:|: |:| :::| |:|| :: |:| |||| |:|| |::|:| :|:|: |:||:|::|| :|::|
Db         154 GSCKRKSKGGSWGKTCRRGARGGSGWSSGAKYKSGSMSKRMWMSSCGRSGCGRRSAYSR 213
```

```
Qy         234 TTCCAGATGGTGGGAAACTGGAGAGAGAGCCCGGTGCACAGGGCGGGGAACCCCGAGGGC 293
                :| | :::| :|:| :|:  :|:| |:  :::: ::|:: | | ||  ||::|:|| |:|
Db         214 YYGTSRKYGTYKKWTYYSASRCMRAYMTTSYSWACSSYTWCRSKRRSMWWKMWRKMRWSR 273
```

```
Qy    294  TGTAGGCCCGTGCCACTGCATGCCTCAGCGCCTGTGTCCTGGCAGCCACCAGCCCCTACT  353
                 :::  ::  :::   :::   ::   ::::: :::::
Db    274  SYGWYSWSYKMWMCTAYKKSYYSRWCYMYRGGGWRGATRYWGRGYMSRVAMMYKKMYWYR  333

Qy    354  GCTGACGGCAGCAGGAATCTGAGCCCGGGAAGGGTTCCAGGGAAGTTCGTGAACCATCTAG  413
                :::::   ::::   :::::: :::
Db    334  GYKGMKRGWWAGRMMWRSMCRWSKACYYMRWMWRMTRRRWAKKSSRTSRRKKRRKWCMR  393

Qy    414  CAAGTCGGGCGTGGGGTGTGGCCAAGTTAGCACACAGATGTAGGGCCCCTTGTGGACTCAGAAA  473
                :::  ::::::::   ::::
Db    394  K-RKYKRMRGYSRMRSCKRARWMKRCRSGRAWKMGCRGGCMTCRMKSYGMWRWKSWKRMAS  452

Qy    474  TTGGCAGCTCTTTTCGGCCCAGGAGGGGGCCCACGGCTGTGTCCGCGGCGTAGGGGTAGCTCAGAAGG  533
                :::   ::::   ::::::   ::::
Db    453  KYKWMSRMYRWRKKKCCSRTTMWGKTRGGWMGCRYKRSGMKRKCRRRWGRMYRMR  512

Qy    534  GTCACCTGGGGGTCTTCCACTACACCCCCGCCTGGACACTGCTGTAGCCCCGAGGGCTCGG  593
                :::::   ::::   :::::::
Db    513  WKRYYMSARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKRMYMYKMMWWYK  572

Qy    594  AGGGACCAGCTGGGACCCCATGAGGAGGAGGAGGCCCAGTTTCTCCCTGTAAGGGTATTCCTGTA  653
                :::    :::::  :::::::  ::::::
Db    573  RKYSKCCSWYCKMSYYASCMKSARKAGGAKMCKRSROMSAWSKSMRSSRKCRKCASKRSSAKR  632

Qy    654  GCATGAGGAGGAACAGACAAGCACCCCAGGGGGGGGACTAACCCGAGGAGATCCAGCCCCCGGGCCCTCACTC  713
                :::::   ::::::    :::::  ::::
Db    633  YAAMCGMTSGSSRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRYRYRAWCM  692

Qy    714  CCGTGTGGGCTCACGGCAAATATCCTAACCTCTCTCTGAGCCCTCCTGCCCAGC  764
                :::  ::
Db    693  YMWRWYYYRYRSYMTYMAWTSSTRMAMTGMKYSGRYWTSWYKYCKCCSWKY  743
```

PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249292P.
PR 17-NOV-2000; 2000US-0249300P.
PR 17-NOV-2000; 2000US-0249302P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250161P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251556P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251999P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457726/49.
DR P-PSDB; AAU20208.
XX
PT Isolated polypeptide for treating, preventing and prognosing disorders
PT related to the endocrine system including endocrine disorders,
PT reproductive disorders, and gastrointestinal disorders and also for
PT testing and detection e.g. diagnosis.
XX
PS Claim 1; SEQ ID NO 71; 558pp; English.
XX
CC The invention relates to cDNAs encoding novel human endocrine antigens or
CC a fragment having biological activity, a domain, an epitope, full length
CC protein, variant, allelic variant or a species homologue of the
CC cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered (e.g. by
CC gene therapy or antisense-therapy). Identifying mutations in the genes
CC coding for the antigens is useful for diagnosing a pathological condition
CC or a susceptibility to a pathological condition. The DNAs, antigens and
CC antibodies raised against the antigens useful for treating, preventing

CC and/ or prognosing disorders related to the endocrine system or hormone
CC imbalance or reproductive disorders, cancers of endocrine tissues,
CC disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC of diseases and disorders are given in the specification. The present
CC sequence encodes an endocrine antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 414 BP; 88 A; 126 C; 105 G; 87 T; 0 U; 8 Other;

Query Match   4.7%; Score 49.2; DB 4; Length 414;
Best Local Similarity   64.3%; Pred. No. 0.011;
Matches   72; Conservative   1; Mismatches   39; Indels   0; Gaps   0;

Qy   67 TCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAAATCTGTCTGAGGCCAC 126
Db   205 TGCCCTTTTAGAGATGAAGGACTGAGGCTCAGAGAAGCCAAGTAATTCACYTAAGGCCAC 146

Qy   127 ACAGCTAGAAAAGCAGCCAGGCCCAGCCAAACCCTGGTGTGTGTCCAGCCCCCA 178
Db   145 GTAGCTAGCAAGAAGCGGAGCCAGGAGACTCAACCTAAGTCTAAGCTGAGCTCCTGCCA 94

RESULT 11
ABN89472/c
ID ABN89472 standard; DNA; 7353 BP.
XX
AC ABN89472;
XX
DT 03-SEP-2002 (first entry)
XX
DE Human G-protein coupled receptor genomic DNA SEQ ID NO:3.
XX
KW Human; G-protein coupled receptor; GPCR; receptor; gene; SNP;
KW single nucleotide polymorphism; chromosome 1; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FH
FT variation   replace(582,G)
FT   /*tag= a
FT   /standard_name= "single nucleotide polymorphism (SNP)"
FT variation   replace(2182,T)
FT   /*tag= b
FT   /standard_name= "single nucleotide polymorphism (SNP)"
FT CDS   3000..4355
FT   /*tag= c
FT   /product= "G-protein coupled receptor"
FT variation   replace(4760,C)
FT   /*tag= d
FT   /standard_name= "single nucleotide polymorphism (SNP)"
XX
PN WO200250272-A2.
XX
PD 27-JUN-2002.
XX
PF 19-DEC-2001; 2001WO-US048547.
XX
PR 20-DEC-2000; 2000US-00740033.
XX
PA (PEKE) PE CORP NY.
XX
PI Gan W, Di Francesco V, Beasley EM;
XX
DR WPI; 2002-500628/53.
DR P-PSDB; ABB81539.
XX
PT Novel isolated G-protein coupled receptor peptide useful for treating
PT disorder characterized by absence of, inappropriate or unwanted
PT expression of the receptor protein, and as immunogens to raise

PT   antibodies.
XX
PS   Claim 4; Fig 3A-C; 72pp; English.
XX
CC   The present sequence encodes a human G-protein coupled receptor (GPCR)
CC   protein (I). (I) can be used in gene therapy. (I) can be used for
CC   identifying a modulator of (I) by contacting (I) with an agent and
CC   determining if the agent has modulated the function or activity of (I).
CC   (I) is also useful for identifying an agent that binds to (I), by
CC   contacting (I) with an agent and assaying the contacted mixture to
CC   determine whether a complex is formed with the agent bound to (I). (I)
CC   can be used in the treatment of a disorder characterised by the absence
CC   of, inappropriate or unwanted expression of the receptor protein. (I) is
CC   located to human chromosome 1
XX
SQ   Sequence 7353 BP; 1772 A; 1742 C; 1920 G; 1919 T; 0 U; 0 Other;
XX

Query Match   4.7%; Score 49.2; DB 6; Length 7353;
Best Local Similarity   59.2%; Pred. No. 0.025; DB 6;
Matches   84; Conservative   0; Mismatches   58; Indels   0; Gaps   0;

Qy   44 ATGTGGTTAGGAGGGCATATTCAGTTCCTCATTTTTCAGATGAGGAGTTGAGGGCCCAGAAC 103
           ||||| ||||| ||| | | |||||||||||||||||||||||||||||||||||||||||
Db   2225 ATGAAGTAGATAATATTTTTATCTCCATTTTGCAAGAGATGAAGCTGAGGGTCTAGAAAGG 2166

Qy   104 GTAAGTAAATCTGTCTGAGGCCACACAGCTAGAAAGCAGCAGAGATGAAGCTGAGGGTCTG 163
           ||||||| ||||| |||| |||||||||||||||||||||||||||||||||||||||||
Db   2165 TTAAGTAACTTACCACAAGGTGCACAGCTACTAAGCGGGCAAGGGGTCTGGGCACCTACCTGG 2106

Qy   164 TGTGTGCAGCCCCCAGCCCAGT 185
           ||| ||||||||||||||||
Db   2105 GCAGTCTGGGCCCCCAGGTCCAGT 2084

RESULT 12
ID   ADA71938/c
      ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO200300898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI   Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR   WPI; 2003-17529O/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1

CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction,
CC   useful for conferring resistance or compatible interaction. (M1) is
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
XX

Query Match   4.7%; Score 48.8; DB 8; Length 2000;
Best Local Similarity   7.8%; Pred. No. 0.021; DB 8;
Matches   48; Conservative   303; Mismatches   265; Indels   2; Gaps   2;

Qy   421 GGCTGGGGTGTGTGGGCCAAGTTAGACACAGATGTAGGSCCCTGTGGACTCAGAAATTGCAG 480
           ::::: ::::: ::::: ::::::::::::
Db   709 KRAKRSYRYRRIRWRWRWKRKGWTYRYRTWRSCRMTRARMSKRRKWAGASMKSCWMYWRGARS 650

Qy   481 CTCTTTTGGCCCCAGGAGGGCCACGCTGTTGSCGGGGGCCTAGCTCAGRAAGGGTCACCT 540
           ::::: ::::::::: :::
Db   649 MWYSKYSCSAKCCKKTRYMYMSSYMSTGMYGMWYSSYXKSMSWTWSKMSYMGKMTCTMYTSMKG 590

Qy   541 GGGGGTCTTCCACTACACCCCCCGCCTGGSACACTGCTGTAGCCCACCCAGGGCTCGGAGGAC 600
           ::::::: ::::: :::: :::
Db   589 STRRSKMGRWSGSMSRMYMRWRWKKKMRKRKYMRYMKWKCTWRRCMCYRKGYTMYTTSRSRMM 530

Qy   601 AGCTGGAGCCCCATGAGGAGAGAGGGGCCAGTTCTCTCCTGTAACGGGGTATTGCTGACATGAG 660
           ::::::::: :::
Db   529 YTGRYKARYTS-KRRRYMWYKXRRKYCWYYYYGMYMKCSYMWRYGYCRACKKCCYAMCWKAA 471

Qy   661 GGAACAGAACAAGGCCCCAGGGGGGACTRAACCCGAGAGATCCAGCCCCGGGGCCTCACTCCCGTG 720
           ::::: ::: ::: :::: ::::::::
Db   529 YTGRYKARYTS-KRRRYMWYKXRRKYCWYYYYGMYMKCSYMWRYGYCRACKKCCYAMCWKAA 471

Qy   721 GCTCACGGCAATATCCTAACCTCTCTGACCTGGGACAGGGGTCTCAGGAAGAGCCTCCTCCTCC 780
           ::::::: :::
Db   470 YSGMMYWYRKYSKWMRMSTKYMWSMWYKYKCRSMKYGAKGCYGCKWMYTCYSGYMKWYTY 411

Qy   781 AGGGGGGTGAGGAAGCCCCAGACACGTGGAAGCCCTTTTTAACCATTCTCGGGGGTGAGGGAGC 840
           ::::::: ::::
Db   410 MGSYKYSRCYKXYM-RMYMYKGNMYMYYSAXYSSMWTWYYYAKYWKYWYKRRGTMSWYGK 352

Qy   841 CCCTTCCCAAATGCCTGGGGTGTCACTGCACTGCTGTGTGGTAGGGGGTCCCCAACGGGGCTC 900
           ::::: :::: ::
Db   351 SYKKKCYCTWMCYMKCMRCYRWRKWMRKKTKYSKRCYCWRYATCYWCCCYRKRGWYSRSRSM 292

Qy   901 AGTGTGGGGCTCGAGGCTGGGCTCTTGGGACAGGGGGTGCTCAGGGAAGAGAGCCTCCTCCTCCC 960
           ::::: :::
Db   291 MRTAGKWKMRSWSRWCRSYSWYKMYXKMWKKSYYMSYGWARSSGTWSRSAAKRTYKGYST 232

Qy   961 TGCCCACTGGGGCATAGGGCCTTCTGGGGAGCTGGGCAGCAATCGTGAGTCAGTGCACCTGG 1020
           ::: :::: :::
Db   231 SRRAKMWRACRMYSACRRYSRTSYYCGCSYCGGSSKWKYMSKSCSMRMTCSSWCSCYTCY 172

Qy   1021 CCCTTCCCGCCCAGCGCGCAG 1038
           ::::: :::: :::: :::
Db   171 YGAMCWSCCCMSMMYMGSCGCYTRGWWKRSKYSMCCKKYCSCCTKYCSYTGYYRYCKWYKY 112

RESULT 13
ID   AAL05121/c
      AAL05121 standard; DNA; 29163 BP.
XX
AC   AAL05121;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 7809.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX

```
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0191123P.
PR   19-MAY-2000; 2000US-0198123P.
PR   07-JUN-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0174487P.
PR   11-JUL-2000; 2000US-0174496P.
PR   14-JUL-2000; 2000US-0182890P.
PR   26-JUL-2000; 2000US-0209663P.
PR   14-AUG-2000; 2000US-0209664P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0252124P.
PR   14-AUG-2000; 2000US-0252213P.
PR   14-AUG-2000; 2000US-0252266P.
PR   14-AUG-2000; 2000US-0252267P.
PR   14-AUG-2000; 2000US-0252268P.
PR   14-AUG-2000; 2000US-0252270P.
PR   14-AUG-2000; 2000US-0254477P.
PR   14-AUG-2000; 2000US-0255759P.
PR   14-AUG-2000; 2000US-0257577P.
PR   18-AUG-2000; 2000US-0257580P.
PR   22-AUG-2000; 2000US-0266681P.
PR   22-AUG-2000; 2000US-0266868P.
PR   23-AUG-2000; 2000US-0271182P.
PR   30-AUG-2000; 2000US-0270009P.
PR   01-SEP-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0292877P.
PR   01-SEP-2000; 2000US-0292943P.
PR   01-SEP-2000; 2000US-0293443P.
PR   05-SEP-2000; 2000US-0293445P.
PR   05-SEP-2000; 2000US-0295090P.
PR   06-SEP-2000; 2000US-0295130P.
PR   06-SEP-2000; 2000US-0304377P.
PR   08-SEP-2000; 2000US-0304388P.
PR   08-SEP-2000; 2000US-0304430P.
PR   08-SEP-2000; 2000US-0312442P.
PR   08-SEP-2000; 2000US-0312443P.
PR   08-SEP-2000; 2000US-0312444P.
PR   08-SEP-2000; 2000US-0314141P.
PR   08-SEP-2000; 2000US-0314142P.
PR   12-SEP-2000; 2000US-0319680P.
PR   14-SEP-2000; 2000US-0321968P.
PR   14-SEP-2000; 2000US-0323397P.
PR   14-SEP-2000; 2000US-0323398P.
PR   14-SEP-2000; 2000US-0323399P.
PR   14-SEP-2000; 2000US-0332400P.
PR   14-SEP-2000; 2000US-0332401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   21-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   14-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0254484P.

PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   29-SEP-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0241826P.
PR   08-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   01-DEC-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
```

```
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 7809; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 29163 BP; 6801 A; 7298 C; 7317 G; 7747 T; 0 U; 0 Other;

Query Match    4.6%;  Score 48.2;  DB 4;  Length 29163;
Best Local Similarity   67.3%;  Pred. No. 0.066;
Matches  68;  Conservative  0;  Mismatches  33;  Indels  0;  Gaps  0;

Qy     51 AGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTA 110
           ||||||||  ||| ||||||| |||||||||| ||||| |||||||| ||||||| |||
Db   8883 AGGTGAGTGTGTTGTTATTCCCATTTTACAAATGAGGAATTGAGGCTCAGAGAGGTGAAATT 8824

Qy    111 ATCTGTCTGAGGCCACACAGCTTAGAGAAAAGCAGCAGCCCAG 151
           ||| |||  |||||| ||||||||  || |||| ||| ||||| 
Db   8823 ACCAGTCCAAGATTACACAGCTTAATGAGCTCCAGAGTCCAG 8783

RESULT 14
ID   ABL98013/c
     ABL98013 standard; DNA; 29163 BP.
XX
AC   ABL98013;
XX
DT   21-JUN-2002  (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 2665.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001329.
XX
PR   31-JAN-2000; 2000US-0179065P.      PR   14-AUG-2000; 2000US-0224519P.
PR   04-FEB-2000; 2000US-0180628P.      PR   14-AUG-2000; 2000US-0225213P.
PR   24-FEB-2000; 2000US-0184664P.      PR   14-AUG-2000; 2000US-0225214P.
PR   02-MAR-2000; 2000US-0166350P.      PR   14-AUG-2000; 2000US-0225266P.
PR   02-MAR-2000; 2000US-0189874P.      PR   14-AUG-2000; 2000US-0225267P.
PR   16-MAR-2000; 2000US-0190076P.      PR   14-AUG-2000; 2000US-0225268P.
PR   17-MAR-2000; 2000US-0190076P.      PR   14-AUG-2000; 2000US-0225270P.
PR   18-APR-2000; 2000US-0198123P.      PR   14-AUG-2000; 2000US-0225447P.
PR   19-MAY-2000; 2000US-0205515P.      PR   14-AUG-2000; 2000US-0225757P.
PR   07-JUN-2000; 2000US-0209467P.      PR   14-AUG-2000; 2000US-0225758P.
PR   28-JUN-2000; 2000US-0214886P.      PR   17-AUG-2000; 2000US-0225759P.
PR   30-JUN-2000; 2000US-0215135P.      PR   18-AUG-2000; 2000US-0226279P.
PR   07-JUL-2000; 2000US-0216647P.      PR   22-AUG-2000; 2000US-0226681P.
PR   07-JUL-2000; 2000US-0216880P.      PR   22-AUG-2000; 2000US-0226686P.
PR   11-JUL-2000; 2000US-0217487P.      PR   22-AUG-2000; 2000US-0227182P.
PR   11-JUL-2000; 2000US-0217496P.      PR   23-AUG-2000; 2000US-0227009P.
PR   14-JUL-2000; 2000US-0218290P.      PR   30-AUG-2000; 2000US-0228224P.
PR   26-JUL-2000; 2000US-0218290P.      PR   01-SEP-2000; 2000US-0229287P.
PR   26-JUL-2000; 2000US-0220963P.      PR   01-SEP-2000; 2000US-0229343P.
PR   26-JUL-2000; 2000US-0220964P.      PR   01-SEP-2000; 2000US-0229344P.
PR   14-AUG-2000; 2000US-0224518P.      PR   01-SEP-2000; 2000US-0229345P.
                                         PR   05-SEP-2000; 2000US-0229509P.
                                         PR   05-SEP-2000; 2000US-0229513P.
                                         PR   06-SEP-2000; 2000US-0230437P.
                                         PR   06-SEP-2000; 2000US-0230438P.
                                         PR   08-SEP-2000; 2000US-0231242P.
                                         PR   08-SEP-2000; 2000US-0231243P.
                                         PR   08-SEP-2000; 2000US-0231244P.
                                         PR   08-SEP-2000; 2000US-0231413P.
                                         PR   08-SEP-2000; 2000US-0231414P.
                                         PR   08-SEP-2000; 2000US-0232080P.
                                         PR   08-SEP-2000; 2000US-0232081P.
                                         PR   12-SEP-2000; 2000US-0231968P.
                                         PR   14-SEP-2000; 2000US-0232397P.
                                         PR   14-SEP-2000; 2000US-0232398P.
                                         PR   14-SEP-2000; 2000US-0232399P.
                                         PR   14-SEP-2000; 2000US-0232400P.
                                         PR   14-SEP-2000; 2000US-0232401P.
                                         PR   14-SEP-2000; 2000US-0233063P.
                                         PR   14-SEP-2000; 2000US-0233064P.
                                         PR   14-SEP-2000; 2000US-0233065P.
                                         PR   21-SEP-2000; 2000US-0234223P.
                                         PR   21-SEP-2000; 2000US-0234274P.
                                         PR   25-SEP-2000; 2000US-0234997P.
                                         PR   25-SEP-2000; 2000US-0234998P.
                                         PR   26-SEP-2000; 2000US-0235484P.
                                         PR   27-SEP-2000; 2000US-0235834P.
                                         PR   27-SEP-2000; 2000US-0235836P.
                                         PR   29-SEP-2000; 2000US-0236327P.
                                         PR   29-SEP-2000; 2000US-0236367P.
                                         PR   29-SEP-2000; 2000US-0236368P.
                                         PR   29-SEP-2000; 2000US-0236369P.
                                         PR   02-OCT-2000; 2000US-0236370P.
                                         PR   02-OCT-2000; 2000US-0236802P.
                                         PR   02-OCT-2000; 2000US-0270377P.
                                         PR   02-OCT-2000; 2000US-0270380P.
                                         PR   02-OCT-2000; 2000US-0270399P.
                                         PR   02-OCT-2000; 2000US-0270400P.
                                         PR   13-OCT-2000; 2000US-0239935P.
                                         PR   13-OCT-2000; 2000US-0239937P.
                                         PR   20-OCT-2000; 2000US-0240960P.
                                         PR   20-OCT-2000; 2000US-0241221P.
                                         PR   20-OCT-2000; 2000US-0241785P.
                                         PR   20-OCT-2000; 2000US-0241786P.
                                         PR   20-OCT-2000; 2000US-0241787P.
                                         PR   20-OCT-2000; 2000US-0241808P.
                                         PR   20-OCT-2000; 2000US-0241809P.
                                         PR   20-OCT-2000; 2000US-0241826P.
                                         PR   01-NOV-2000; 2000US-0244177P.
                                         PR   01-NOV-2000; 2000US-0244617P.
                                         PR   08-NOV-2000; 2000US-0246474P.
                                         PR   08-NOV-2000; 2000US-0246475P.
                                         PR   08-NOV-2000; 2000US-0246476P.
                                         PR   08-NOV-2000; 2000US-0246477P.
```

PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0249208P.
PR   08-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   01-DEC-2000; 2000US-0249216P.
PR   05-DEC-2000; 2000US-0249217P.
PR   05-DEC-2000; 2000US-0249218P.
PR   06-DEC-2000; 2000US-0249244P.
PR   06-DEC-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483232/52.
XX
PT   Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT   for preventing, diagnosing and/or treating testicular cancer.
XX
PS   Disclosure; SEQ ID NO 2665; 766pp; English.
XX
CC   The present invention provides the protein and coding sequences of 973
CC   human testicular antigens, and fragments of their genomic sequences. The
CC   sequences can be used in the treatment of cardiovascular, urinary system,
CC   reproductive system, immune, respiratory, neurological and
CC   gastrointestinal disorders, infections, and particularly cancer,
CC   especially testicular cancers. The present sequence is a DNA encoding a
CC   protein fragment of the invention.
XX
SQ   Sequence 29163 BP; 6801 A; 7298 C; 7317 G; 7747 T; 0 U; 0 Other;

Qy   51 AGGAGGGCATATTCAGTCCCATTTTTTCAGATGAGGAGTTCAGGCCCAGAGAGAACGTAAGTA 110
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8883 AGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTCAGGCCCAGAGAGAACGTAAGTT 8824

Qy   111 ATCTGTGTTGAGGCCACACAGCTTAGAAAGCAGCCAGGCCCAG 151
Query Match   4.6%; Score 48.2; DB 4; Length 29163;
Best Local Similarity   67.3%; Pred. No. 0.066;
Matches   68; Conservative   0; Mismatches 33; Indels   0; Gaps   0;

Qy   111 ATCTGTGTTGAGGCCACACAGCTTAGAAAGCAGCCAGGCCCAG 151
            |||||||||||||||||||||||||||||||||||||||||||
Db   8823 ACCAGTCCAAGATTACACAGCTTAATGAGCTGCAGAGTCCAG 8783

Db   8823 ACCAGTCCAAGATTACACAGCTTAATGAGCTGCAGAGTCCAG 8783

RESULT 15
ADA02876
ID   ADA02876 standard; DNA; 52661 BP.
XX
AC   ADA02876;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Human DPT carcinoma associated gene, SEQ ID NO:1394.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FF   17-JUL-2003.
PD
PN   WO200305714 6-A2.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1394; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   oncogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   pathogenic (especially breast cancer, prostate cancer, lymphoma or
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 52661 BP; 15800 A; 11001 C; 11082 G; 14778 T; 0 U; 0 Other;

Qy   18 CAGTCCTCCACACTCCAGTCCCACAGATGTGGTTAGGGACGGGCATATTCAGTCCCATTTTTTC 77
                ||||| |||||||||||||| |||| | ||||||||||||||| |||||||||||||||||||
Db   49680 CAGGCTTTCCCATTTGCTACTCGGCAGATACCCTAGACTCGGCAGTGTTACTGCCATTTTAT 49739

Query Match   4.5%; Score 47.2; DB 9; Length 52661;
Best Local Similarity   55.5%; Pred. No. 0.14;
Matches   91; Conservative   0; Mismatches 73; Indels   0; Gaps   0;

```
Qy      78 AGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCACACAGCTAGAAA 137
           |||||   ||| | |||  ||||  ||| || |||| ||  | |||  |||||| ||||||
Db   49740 AGATAGGTTCAAAGGTCCCCAAAATAATAGGTTGACTTGCCCAAAGGTCACACAGCTAGTGG 49799

Qy     138 GCAGCCAGGCCGCACCCGAACCCCTGGTGTGTGCAGCCCCCAGCC 181
           || ||||  | |  | |  |||  ||| | | ||| ||||| ||||
Db   49800 AAATGGATTGAACTGAAGTCTAGCTGTTTACAAAGCCCAGTC 49843
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 15:14:58 ; Search time 219 Seconds
               (without alignments)
               7755.503 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 caggaggaccggaagttgcag..........ggccttcccggcaggcgcag 1038

Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81818359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Match	Length	DB	ID	Description
c 1	55.8	5.4		72455	4	US-09-949-016-13793	Sequence 13793, A
c 2	55.8	5.4		104475	4	US-09-949-016-12115	Sequence 12115, A
c 3	55.8	5.4		111282	3	US-09-754-250-3	Sequence 3, Appli
c 4	50	4.8		255171	4	US-09-949-016-12822	Sequence 12822, A
c 5	50	4.8		256176	4	US-09-949-016-15524	Sequence 15524, A
c 6	47.8	4.6		24817	4	US-09-949-016-16177	Sequence 16177, A
c 7	47.2	4.5		473	4	US-09-621-976-14118	Sequence 14118, A
c 8	47	4.5		19437	4	US-09-949-016-17032	Sequence 17032, A
c 9	47	4.5		19437	4	US-09-949-016-17033	Sequence 17033, A
c 10	46.8	4.5		601	4	US-09-949-016-18953	Sequence 18953, A
c 11	46.8	4.5		601	4	US-09-949-016-185108	Sequence 185108,
c 12	46.8	4.5		43562	4	US-09-949-016-16222	Sequence 16222, A
c 13	46.2	4.5		703	3	US-09-313-300-6	Sequence 6, Appli
c 14	46.2	4.5		57837	4	US-09-949-016-14371	Sequence 14371, A
c 15	46.2	4.5		57839	4	US-09-949-016-17601	Sequence 17601, A
c 16	45.4	4.4		601	4	US-09-949-016-184954	Sequence 184954,
c 17	45.4	4.4		601	4	US-09-949-016-185109	Sequence 185109,
c 18	45	4.3		601	4	US-09-949-016-183463	Sequence 183463,
c 19	44.8	4.3		601	4	US-09-949-016-30076	Sequence 30076,
c 20	44.8	4.3		601	4	US-09-949-016-72407	Sequence 72407, A
c 21	44.8	4.3		601	4	US-09-949-016-92404	Sequence 92404, A
c 22	44.8	4.3		601	4	US-09-949-016-92405	Sequence 92405, A
c 23	44.8	4.3		601	4	US-09-949-016-9206	Sequence 92406, A
c 24	44.8	4.3		601	4	US-09-949-016-137895	Sequence 137895,
c 25	44.8	4.3		601	4	US-09-949-016-137896	Sequence 137896,
c 26	44.8	4.3		601	4	US-09-949-016-137897	Sequence 137897,
c 27	44.8	4.3		52711	4	US-09-949-016-12224	Sequence 12224, A
c 28	44.8	4.3		52865	4	US-09-949-016-15618	Sequence 15618, A
c 29	44.4	4.3		72602	4	US-09-949-016-14385	Sequence 14385, A
c 30	44.4	4.3		30006	4	US-09-949-016-16385	Sequence 16385, A
c 31	44.4	4.3		57299	4	US-09-949-016-12141	Sequence 12141, A
c 32	44.4	4.3		300598	4	US-09-949-016-11868	Sequence 11868, A
c 33	44.4	4.3		302604	4	US-09-949-016-14588	Sequence 14588, A
c 34	44.4	4.3		302604	4	US-09-949-016-14589	Sequence 14589, A
c 35	44.4	4.3		308362	4	US-09-949-016-17119	Sequence 17119, A
c 36	43.8	4.2		601	4	US-09-949-016-183464	Sequence 183464,
c 37	43.8	4.2		36203	4	US-09-949-016-15430	Sequence 15430, A
c 38	43.6	4.2		46118	4	US-09-949-016-17003	Sequence 17003, A
c 39	43.6	4.2		13995	4	US-09-949-016-14231	Sequence 14231, A
c 40	43.6	4.2		36387	4	US-09-949-016-12370	Sequence 12370, A
c 41	43.6	4.2		36387	4	US-09-949-016-13862	Sequence 13862, A
c 42	43.6	4.1		601	4	US-09-949-016-92407	Sequence 92407, A
c 43	42.6	4.1		601	4	US-09-949-016-137898	Sequence 137898, A
c 44	42.6	4.1		312474	4	US-09-949-016-17434	Sequence 17434, A
c 45	42	4.0		37004	4	US-09-949-016-15317	Sequence 15317, A

ALIGNMENTS

RESULT 1
US-09-949-016-13793/c
Sequence 13793, Application US/09949016
; Patent No. 6812339
GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13793
; LENGTH: 72455
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(72455)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13793

Query Match   5.4%; Score 55.8; DB 4; Length 72455;
Best Local Similarity   70.1%;
Matches   89; Conservative   0; Mismatches 37; Indels   1; Gaps   1;

```
Qy        67 TCCCATTTTCAGATGAGGAGGACTTGAGGCCCAGAGAGACGTAAGTAATCTGTCTGAGGCCAC 126
             |||||||| |||||| ||| |||   ||||||| || |||||||||  |  |||| |||
Db     55484 TCCCCATTTTCAGGTGTGAAACTGAGGCGCCAGAGAGGTTTAGTGACTTGCTCAAGGTCAC 55425

Qy       127 ACAGCTAGAAAGCAGCAGCCAGGCC-CAGCCGGAACCCTGGGTGTTGTGCAGCCCCCAGCCCCAGT 185
             ||||| |||||| || |||  |||| ||||| || ||||||  ||||| |||| ||||| |||
Db     55424 ACAGCTCAAAAGCAGCAAAAGCCGAACCTGAAACCCAGGTCGTCGTCCAGCTCCCTGGCCAGG 55365

Qy       186 TGCTCAT 192
             || |||
Db     55364 CCATCTT 55358
```

RESULT 2
US-09-949-016-12115/c
Sequence 12115, Application US/09949016
; Patent No. 6812339

```
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION:   WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12115
;   LENGTH: 104475
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(104475)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12115

Query Match             70.1%;  Score 55.8;  DB 4;  Length 104475;
Best Local Similarity   70.1%;  Pred. No. 5.4e-05;
  Matches   89;  Conservative    0;  Mismatches  37;  Indels    1;  Gaps    1;

Qy        67 TCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCAC 126
             |||| ||||||||| ||||||||||||||||||||| ||||||||| || |||||||||||
Db     87505 TCCCATTTTTCAGATGAGGAAACTGAGGCCCAGAGAGGAGGTTTAGTGACTTGAGGTCAC 87446

Qy       127 ACAGCTAGAAAGAGCAGCCAGGCC--CAGCCGAGAACCCCTGGTTGTGTGTGCAGCCCAGT 185
             ||||| |||||||||||||||| ||  |||| |||||| |||||||| |||||  |||||
Db     87445 ACAGCTCAAAAGCAGCAAAGCCGAAGCTGAAACCCAGGTCTGTCCCAGCTCCCTGGCCAGG 87386

Qy       186 TGCTCAT 192
             | |||||
Db     87385 CCATCTT 87379
```

```
RESULT 3
US-09-754-250-3/c
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
;   APPLICANT: WEI, Ming-Hui et al
;   TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
;   TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
;   TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
;   FILE REFERENCE: CL001063
;   CURRENT APPLICATION NUMBER: US/09/754,250
;   CURRENT FILING DATE: 2001-01-05
;   NUMBER OF SEQ ID NOS: 5
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 111282
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(111282)
;   OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3

Query Match              5.4%;  Score 55.8;  DB 3;  Length 111282;
Best Local Similarity   70.1%;  Pred. No. 5.5e-05;
  Matches   89;  Conservative    0;  Mismatches  37;  Indels    1;  Gaps    1;

Qy        67 TCCCATTTTTCAGATGTGTGAAAACTGAGGCCCAGAGAGGAGGTTTAGTGACTTGAGGTCAC 126
             |||| ||||||||||| | ||||||||||||||||||||||||||||| ||||||||||||
Db     93162 TCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCAC 93103
```

```
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(256176)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15524

Query Match   4.8%;  Score 50;  DB 4;  Length 256176;
Best Local Similarity  72.2%;
Matches  65;  Conservative  0;  Mismatches  25;  Indels  0;  Gaps  0;

Qy     59 ATATTCAGTTCCCATTTTTCAGATGAGGAGTTGAGGCGCCAGAGAACGTAGTAAGTAAATCTGTGTCT 118
Db 189846 ATTGTTAACCCCATTTTACAGATGAGGAAATGAGACAGAGAAGAGATTAAGTAATCCGTTCC 189787

Qy    119 GAGGGCCACACAGCTAGAAAAGCAGACCAGGCC 148
Db 189786 AAGGTCACACAGCCTAGTGAGTGGGCAAAGCC 189757


RESULT 6
US-09-949-016-16177
; Sequence 16177, Application US/0994016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 16177
; LENGTH: 24817
; TYPE: DNA
; ORGANISM: Human
; SOFTWARE: FastSEQ for Windows Version 4.0
US-09-949-016-16177

Query Match   4.6%;  Score 47.8;  DB 4;  Length 24817;
Best Local Similarity  55.8%;  Pred. No. 0.0057;
Matches  91;  Conservative  0;  Mismatches  72;  Indels  0;  Gaps  0;

Qy     29 ACTCAGTTCCCACAGATGTGGTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGT 88
Db   7117 ACTCAGAACATAGACATGAGACAGGTACAATTATTTACCCAAATTATTCATGATGGTGAAC 7176

Qy     89 TGAGGCCCCAGAGAACGTAGTAATCTGTCTGGAGGCCCACACAGCTAGAAAAGCAGCCAGGCC 148
Db   7177 TAAGGCACAGAGACAGTAAGAAAGTTGCCAGAGGTCACACAGCTAATAAGTGGTCGTTCC 7236

Qy    149 CAGCCGAACCCTGGTGTGTGTGCAGCCCCCAGCCCAGTTGCTCA 191
Db   7237 AGGGCATGGGCACCATGCATCCAGCCTCAGTCCATGGTGTGCA 7279


RESULT 7
US-09-621-976-14118/c
; Sequence 14118, Application US/0921976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14118
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-14118

Query Match   4.5%;  Score 47.2;  DB 4;  Length 473;
Best Local Similarity  67.5%;  Pred. No. 0.0016;
Matches  81;  Conservative  0;  Mismatches  38;  Indels  1;  Gaps  1;

Qy     30 CTCAGTTCCCACAGATGTGGTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGG-AGT 88
Db    194 CTCACAACACCCCCTGTGAGGTAGGTCAGTATTATTATCCCCATTTTACAGATGGGGGAAC 135

Qy     89 TGAGGCCCCAGAGAACCTAAGTAAACTCGTTCTGAGGCCCCAGCTAGGAGCCAGCCAGGCC 148
Db    134 TGAGGCACAGAGAGAGGTTAAGTGACTTGCCCAAAGGCCACACACAGCAGCCAGCGCGGTCAAGCC 75


RESULT 8
US-09-949-016-17032
; Sequence 17032, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 17032
; LENGTH: 194937
; TYPE: DNA
; ORGANISM: Human
; SOFTWARE: FastSEQ for Windows Version 4.0
US-09-949-016-17032

Query Match   4.5%;  Score 47;  DB 4;  Length 194937;
Best Local Similarity  58.0%;  Pred. No. 0.023;
Matches  83;  Conservative  0;  Mismatches  60;  Indels  0;  Gaps  0;

Qy      1 CAGGGACTCGAAGTTGCAGTCCTCCACACTCAGTTCCCACAGATGTGGTAGGAGGGCAT 60
Db 191265 CAGGTCAGGGAGGCTTATATCCCCTGCGTCTTTCACACTCACAGCAAACCCTGCAGAGTGGGGAGT 191324

Qy     61 ATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGGCCCCAGAGAACGTAAGTAATCTGTCTGA 120
Db 191325 GACTATTCCCATTCTTTAGATGAGGAAATCAAAGCTCAGAGAGGTTAAATAACTTGCCTAA 191384

Qy    121 GGCCACACAGCTAGAAAAGCAGCC 143
Db 191385 GGCCACACAGCAGGACTGCAGC 191407


RESULT 9
US-09-949-016-17033
; Sequence 17033, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED, METHODS OF DETECTION AND USES THEREOF
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17033
;   LENGTH: 194937
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17033

Query Match             58.0%;  Score 47;  DB 4;  Length 194937;
Best Local Similarity   4.5%;
Matches   83;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy           1 CAGGGACTCCGAAGTTCCAGTTCCACACTCAGTTCCTCCACACTCAGTTCCCACACACAT 60
                 |||||||  ||   |||||   |||   ||| |   |||   |||   ||| 
Db      191265 CAGGTGCAGGGAGCTTATATCCCCTGCTTTCACACTCACACGAACCCTCCAGTAGGAGT 191324

Qy          61 ATTCAGTCCCATTTTTCAGATGAGGAGGAGTTGAGCCCCAGAGAACGTAAGTAAACTGTGTCTGA 120
                 ||   |||   ||||   |||   ||   ||| 
Db      191325 GACTATTCCCATTCTTTAGATGAGGAATCAAAGCTCAGAGAGGTTAAATAACTTGCCTAA 191384

Qy         121 GGCCACACAGCTAGAAAGCAGCC 143
                 |||||||||||||   ||| 
Db      191385 GGGCACACAGCAGGAACTGCGAGGC 191407
```

RESULT 10
US-09-949-016-184953
Sequence 184953, Application US/0949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 184953
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-184953

Query Match   61.5%; Score 46.8; DB 4; Length 601;
Best Local Similarity   4.5%;
Matches   75; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

```
Qy          22 CCTCCACACTCAGTTCCCACACAGATGTGGTTAGGAGGGGGCATATTCAGTCCCATTTTTCAGAT 81
                 |||   ||||||  ||   |||   ||| 
Db          19 CCTGCTTTCACACTCACACGAACCCTGCAGTAGGAGGTGA 78

Qy          82 GAGGAGTTGAGGGCCCAGAGAACGTAAGTAAATCTGTCTCTGAGGGCCACACAGCTAGAAAGCAG 141
                 |||||||||||||||   ||| 
Db          79 GAGGAATCAAAGCTCAGAGAGGTTAAATAACTTGCCTAAGGCCACACAGCAGGAGACTGCAG 138

Qy         142 CC 143
                 ||
Db         142 CC 143
```

Db   139 GC 140

RESULT 11
US-09-949-016-185108
Sequence 185108, Application US/0949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 185108
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-185108

Query Match   61.5%; Score 46.8; DB 4; Length 601;
Best Local Similarity   4.5%;
Matches   75; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

```
Qy          22 CCTCCACACTCAGTTCCCACACAGATGTGGTTAGGAGGGGGCATATTCAGTCCCATTTTTCAGAT 81
                 |||   ||||||  ||   |||   ||| 
Db          19 CCTGCTTTCACACTCACACGAACCCTGCAGTAGGAGGTGA 78

Qy          82 GAGGAGTTGAGGGCCCAGAGAACGTAAGTAAATCTGTCTCTGAGGGCCACACAGCTAGAAAGCAG 141
                 |||||||||||||||   ||| 
Db          79 GAGGAATCAAAGCTCAGAGAGGTTAAATAACTTGCCTAAGGCCACACAGCAGGAGACTGCAG 138

Qy         142 CC 143
                 ||
Db         139 GC 140
```

RESULT 12
US-09-949-016-16222/c
Sequence 16222, Application US/0949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 16222
;   LENGTH: 43562
;   TYPE: DNA
;   ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(43562)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16222

RESULT 13
US-09-313-300-6
Sequence 6, Application US/09313300
Patent No. 6222027

GENERAL INFORMATION:
APPLICANT: Kaser, Matthew, R.
APPLICANT: Lal, Preeti
APPLICANT: Yue, Henry
APPLICANT: Tang, Tom Y.
APPLICANT: Baughn, Mariah, R.
APPLICANT: Azimzai, Yalda
TITLE OF INVENTION: MOLECULES EXPRESSED IN HIPPOCAMPUS
FILE REFERENCE: PB-0012 US
CURRENT APPLICATION NUMBER: US/09/313,300
CURRENT FILING DATE: 1999-05-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PERL Program
PUBLICATION INFORMATION: 1941247
OTHER INFORMATION:
SEQ ID NO 6
LENGTH: 703
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY:

US-09-313-300-6

Query Match   63.2%;   Score 46.8;   DB 4;   Length 43562;
Best Local Similarity   63.2%;   Pred. No. 0.014;
Matches   72;   Conservative   0;   Mismatches   42;   Indels   0;   Gaps   0;

Qy   38 CCACAGATGTGGTAGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCA   97
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 38106 CAACAACTCTATGAGATAAAAACAATTATTTCCATTTTACAGTCAAAGAATTGAGGCACA 38047

Qy   98 GAGAACGTAAGTAATCTGTCTGAGGCCCACACAGCTAGAAAGCAGCCAGGCCCAG   151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 38046 GAGATGTTAAGTAATTAATTTACCTAAGGTCACACAACTAATTAATTGGGCAGGGGGCAG 37993

RESULT 14
US-09-949-016-14371
Sequence 14371, Application US/0994016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSEQ for Windows Version 4.0

Query Match   67.2%;   Score 46.2;   DB 3;   Length 703;
Best Local Similarity   67.2%;   Pred. No. 0.0036;
Matches   80;   Conservative   0;   Mismatches   38;   Indels   1;   Gaps   1;

Qy   30 CTCAGTTCCCACAGATGTGGTAGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGG-AGT   88
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   510 CTCACAACACCCCTGTGAGGTAGGTCAGTGATTATTATTATCCCCATTTTTACAGATGGG 569

Qy   89 TGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCCACAGCTAGAAAGCAGCCAGGC   147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570 TGAGGCACAGAGAGAGGTTAAGTGACTTGCCCAAGGCCACACAGCCGAGCCAGGGTCAAGC 628

SEQ ID NO 14371
LENGTH: 57837
TYPE: DNA
ORGANISM: Human
US-09-949-016-14371

Query Match   58.3%;   Score 46.2;   DB 4;   Length 57837;
Best Local Similarity   58.3%;   Pred. No. 0.023;
Matches   81;   Conservative   0;   Mismatches   58;   Indels   0;   Gaps   0;

Qy   31 TCAGTTCCCACAGATGTGGTAGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTG   90
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 20341 TGAATTCTCAGAATAACCTTATGATGTCAGGATCGTGCCAATTTTACACATTAGGAGCAG 20400

Qy   91 AGGGCCCAGAGAACGTAAGTAAGTAATCTGTCTGAGGCCCACACAGCTAGAAAGCAGCCAGCCA   150
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 20401 TGGCACAGAGAAAGATAAAAGTAAATTGACTAAGGCCACACCACGAGTCAGTGGGGGAGGCCTGG 20460

Qy   151 GCCGAACCCCTGGGTGTGTG   169
          ||||||||||||||||||||
Db 20461 ACTGGGCTCTCAGGCATGTG 20479

RESULT 15
US-09-949-016-17601
Sequence 17601, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 17601
LENGTH: 57839
TYPE: DNA
ORGANISM: Human
US-09-949-016-17601

Query Match   58.3%;   Score 46.2;   DB 4;   Length 57839;
Best Local Similarity   58.3%;   Pred. No. 0.023;
Matches   81;   Conservative   0;   Mismatches   58;   Indels   0;   Gaps   0;

Qy   31 TCAGTTCCCACAGATGTGGTAGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTG   90
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 20342 TGAATTCTCAGAATAACCTTATGATGTCAGGATCGTGCCAATTTTACACATTAGGAGCAG 20401

Qy   91 AGGGCCCAGAGAACGTAAGTAAGTAATCTGTCTGAGGCCCACACAGCTAGAAAGCAGCCAGCCA   150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 20402 TGGCACAGAGAAAGATAAAAGTAAATTGACTAAGGCCACACCACGAGTCAGTGGGGGAGGCCTGG 20461

Qy   151 GCCGAACCCCTGGGTGTGTG   169
          ||||||||||||||||||||
Db 20462 ACTGGGCTCTCAGGCATGTG 20480

Search completed: June 29, 2005, 18:13:20
Job time : 223 secs

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 15:32:48 ; Search time 757 Seconds
(without alignments)
8570.723 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 cagggactcgaagttcgag..........ggcccctcccgccagcgcag 1038

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 6067389 seqs, 312525875S residues

Total number of hits satisfying chosen parameters: 12134778

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1038	100.0	3006	17	US-10-189-099A-1	Sequence 1, Appli
2	1038	100.0	3006	21	US-10-617-443B-1	Sequence 1, Appli
3	55.8	5.4	111282	13	US-10-094-989-3	Sequence 3, Appli
4	52.4	5.0	7086	10	US-09-764-891-5712	Sequence 5712, Ap
5	52.4	5.0	7388	10	US-09-764-891-5713	Sequence 5713, Ap
6	49.6	4.8	532	17	US-10-074-024-572	Sequence 572, App
7	49.6	4.8	710	13	US-10-027-632-32570	Sequence 32570, A

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

c	8	49.6	4.8	710	17	US-10-027-632-32570	Sequence 32570, A
c	9	49.2	4.7	414	17	US-10-074-024-71	Sequence 71, Appl
c	10	49.2	4.7	7353	9	US-09-740-033-3	Sequence 3, Appli
c	11	48.2	4.7	29163	10	US-09-764-891-7809	Sequence 7809, Ap
c	12	48	4.6	201	21	US-10-741-600-34521	Sequence 34521, A
c	13	48	4.6	585	13	US-10-027-632-226737	Sequence 226737,
c	14	48	4.6	585	13	US-10-027-632-226738	Sequence 226738,
c	15	48	4.6	585	13	US-10-027-632-226739	Sequence 226739,
c	16	48	4.6	585	13	US-10-027-632-226737	Sequence 226737,
c	17	48	4.6	585	13	US-10-027-632-226738	Sequence 226738,
c	18	48	4.6	585	21	US-10-027-632-226739	Sequence 226739,
c	19	48	4.6	274869	21	US-10-741-600-17650	Sequence 17650, A
c	20	47.4	4.6	669	13	US-10-027-632-132988	Sequence 132988,
c	21	47.4	4.6	669	13	US-10-027-632-132988	Sequence 132988,
c	22	47.4	4.6	669	13	US-10-027-632-132855	Sequence 132855,
c	23	47.2	4.5	1238	13	US-10-027-632-122855	Sequence 122855,
c	24	47.2	4.5	1238	13	US-10-027-632-122856	Sequence 122856,
c	25	47.2	4.5	1238	17	US-10-027-632-226739	Sequence 226739,
c	26	47.2	4.5	52661	11	US-10-997-722-142	Sequence 142, App
c	27	47.2	4.5	74822	18	US-10-240-425-1121	Sequence 1121, Ap
c	28	47.2	4.5	99957	11	US-09-997-722-298	Sequence 298, App
c	29	47	4.5	52745	19	US-10-741-601-5726	Sequence 5726, Ap
c	30	47	4.5	253861	19	US-10-741-601-5611	Sequence 5611, Ap
c	31	47	4.5	261817	13	US-10-087-192-2002	Sequence 2002, Ap
c	32	47	4.5	18750	20	US-09-719-993-6900	Sequence 6900, Ap
c	33	46.8	4.5	13216	10	US-09-764-891-7810	Sequence 7810, Ap
c	34	46.4	4.5	562	13	US-10-027-632-127376	Sequence 127376,
c	35	46.4	4.5	562	13	US-10-027-632-127376	Sequence 127376,
c	36	46.2	4.5	201	20	US-10-719-993-42166	Sequence 42166, A
c	37	46.2	4.5	2849	17	US-10-104-047-436	Sequence 436, App
c	38	46.2	4.4	63686	13	US-10-087-192-466	Sequence 466, App
c	39	45.4	4.4	32944	20	US-10-719-993-6978	Sequence 6978, Ap
c	40	45	4.3	527	13	US-10-027-632-105773	Sequence 105773,
c	41	45	4.3	527	13	US-10-027-632-105773	Sequence 105773,
c	42	45	4.3	729	13	US-10-027-632-18523	Sequence 18523, A
c	43	45	4.3	729	17	US-10-027-632-18523	Sequence 18523, A
c	44	45	4.3	653458	21	US-10-461-862-4	Sequence 4, Appli
c	45	44.6	4.3	600	13	US-10-027-632-288733	Sequence 288733,

ALIGNMENTS

RESULT 1
US-10-189-099A-1
Sequence 1, Application US/10189099A
Publication No. US20040005659A1

GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1
; LENGTH: 3006
; TYPE: cDNA
; ORGANISM: Homo sapiens
US-10-189-099A-1

Query Match 100.0%; Score 1038; DB 17; Length 3006;
Best Local Similarity 100.0%;
Matches 1038; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGGACTCGAAGTTCGAGTCCTCCACACTCAGTTCCCACAGATGTGGTAGGAGGGCAT 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CAGGGACTCGAAGTTCGAGTCCTCCACACTCAGTTCCCACAGATGTGGTAGGAGGGCAT 60

Qy 61 ATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAGAACGTAAGTAATCTGTCTGA 120

```
RESULT 2
US-10-617-443B-1
; Sequence 1, Application US/10617443B
; Publication No. US20050019777A1
; GENERAL INFORMATION:
```

;

;

APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; CURRENT FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-617-443B-1

Query Match   100.0%;   Score 1038;   DB 21;   Length 3006;
Best Local Similarity   100.0%;   Pred. No. 2e-305;
Matches 1038;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 CACGGGACTCGAAGTTTGCAGTCTCCACACTCAGTTCCCACACGATGTGTAGGAGGGCAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACGGGACTCGAAGTTTGCAGTCTCCACACTCAGTTCCCACACGATGTGTAGGAGGGCAT 60

Qy     61 ATTCAGTCCCATTTTTCAGATGAGGAGTTGAGCCCAGAGAACGTAAGTAAATCTGTCTGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTCAGTCCCATTTTTCAGATGAGGAGTTGAGCCCAGAGAACGTAAGTAAATCTGTCTGA 120

Qy    121 GGCCACACAGCTAGAAAGCAGCAGCCCAGGCCCAGGCCCAACCCCTGGTGTGTCAGCCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCCACACAGCTAGAAAGCAGCAGCCCAGGCCCAGGCCCAACCCCTGGTGTGTCAGCCCC 180

Qy    181 CCAGTTGCTCATTGCGGGGCTCGGGAGCCCAGCAGCCAGCCAGCCATGTGTTCCAGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGTTGCTCATTGCGGGGCTCGGGAGCCCAGCAGCCAGCCAGCCATGTGTTCCAGA 240

Qy    241 TGGTGGGAACTGGAGAGAGCCCGCACAGGCCCGTGCAGGGAACCCTGTAGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGTGGGAACTGGAGAGAGCCCGCACAGGCCCGTGCAGGGAACCCTGTAGGC 300

Qy    301 CCCGTGCCACTGCATGCCTCAGGCCTGGTCCTGGCAGCCCACAGCCCCTACTGCTTGACG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCCGTGCCACTGCATGCCTCAGGCCTGGTCCTGGCAGCCCACAGCCCCTACTGCTTGACG 360

Qy    361 GCAGCAGGAAATCTGAGCCCGGGAGAGGGTCCAGGGAAGTTCGTGAACAATCTAGCAAGTCG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCAGGAAATCTGAGCCCGGGAGAGGGTCCAGGGAAGTTCGTGAACAATCTAGCAAGTCG 420

Qy    421 GGCTGGGGGTGTGGCCCAGTTAGACACAGATGTAGGGGCCCTGTGACTCAGAGAAATTGGCAG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGCTGGGGGTGTGGCCCAGTTAGACACAGATGTAGGGGCCCTGTGACTCAGAGAAATTGGCAG 480

Qy    481 CTCTTTTGGCCCAGAGGGGCCCAGGGGTTCAGGAGCTCAGAAGGGTCACCT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTCTTTTGGCCCAGAGGGGCCCAGGGGTTCAGGAGCTCAGAAGGGTCACCT 540

Qy    541 GGGGGTCTTCCACTACACCCCCGCCCTGGACACTGCTGTCTGTAGCGTCGTGTTCAGCATGAG 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGGGGTCTTCCACTACACCCCCGCCCTGGACACTGCTGTCTGTAGCGTCGTGTTCAGCATGAG 600

Qy    601 AGCTGGAGCCCATGAGAGGAGAGGCCAGTTCTCTCTCCTGTAAGGGTATTCCTGTAGCATGAG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGCTGGAGCCCATGAGAGGAGAGGCCAGTTCTCTCTCCTGTAAGGGTATTCCTGTAGCATGAG 660

Qy    661 GGAACAGACAAGGCCCAGGGGGACTAACCCGGACCCTCAGCCCTCACTCCGTGTG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGAACAGACAAGGCCCAGGGGGACTAACCCGGACCCTCAGCCCTCACTCCGTGTG 720

Qy    721 GCTCACGGCAATATCCTAACCTCTCTCTGAGCCTCCAGCCTAGCAGGGGTCCAGTG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCTCACGGCAATATCCTAACCTCTCTCTGAGCCTCCAGCCTAGCAGGGGTCCAGTG 780

Qy    781 AGGGGGGTGAGGAAGCCCAGCACGTGGAAGCCTTTTAACCATTCTCGGGGTGAGCGAGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGGGGGGTGAGGAAGCCCAGCACGTGGAAGCCTTTTAACCATTCTCGGGGTGAGCGAGC 840

Qy    841 CCCTTCCCAAAATGCCTGGTGTCACTGCACTGTGTGTGTAGGGGGTAGGGGGTCACCT 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCCTTCCCAAAATGCCTGGTGTCACTGCACTGTGTGTGTAGGGGGTAGGGGGTCACCT 900

Qy    901 AGTGTGGGGCTGAGGCTGGCTCTGAACTGGGACTGGGGGTTCTCAGGACACAGAGAGCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AGTGTGGGGCTGAGGCTGGCTCTGAACTGGGACTGGGGGTTCTCAGGACACAGAGAGCC 960

Qy    961 TGCCCACTGGGCATAGGCCTCTGGGAGCTGGCAGCATGGTGGCAGCATCGTGATCACTGTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGCCCACTGGGCATAGGCCTCTGGGAGCTGGCAGCATGGTGGCAGCATCGTGATCACTGTG 1020

Qy   1021 CCCTTCCCGCCACGGCAG 1038
          ||||||||||||||||||
Db   1021 CCCTTCCCGCCACGGCAG 1038
```

RESULT 3
US-10-094-989-3/c
; Sequence 3, Application US/10094989
; Publication No. US20020115179A1
; GENERAL INFORMATION:
; APPLICANT: WEI; Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL0010063DIV
; CURRENT APPLICATION NUMBER: US/10/094,989
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/754,250
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(111282)
; OTHER INFORMATION: n = A,T,C or G
US-10-094-989-3

Query Match   5.4%;   Score 55.8; DB 13; Length 111282;
Best Local Similarity   70.1%; Pred. No. 3.1e-06;
Matches   89; Conservative   0; Mismatches   37; Indels   1; Gaps   1;

Qy   67 TCCCATTTTTCAGATGAGGACTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGAGGCCAC 126
          |||| ||||||||||||||||| ||||| | ||||| |||| | ||||| ||| ||||||
Db 93162 TCCCCATTTTCAGCGTGTGAAACTGAGGCCCAGAGAGGTTTAGTGACTTGCTCAAGGTCAC 93103

Qy   127 ACAGCTAGAAAAGCAGCCACGGCCC-CAGCCGAACACCCTGGTCGTGTCGACCGCCAGT 185
          ||| ||||| |||| | || |||| |||| |||||| | ||| | |||| ||||
Db 93102 ACAGCTCAAAAGCAGCAAAACCCGAGGCTGAAACCTGAAAACCCAGGTCTGTCGCCCTGGCCAGG 93043

RESULT 4
US-09-764-891-5712
; Sequence 5712, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17

Qy   186 TGCTCAT 192
          ||||||
Db 93042 CCATCTT 93036

Qy   1021 CCCTTCCCGCCAGCGCAG 1038
          ||||||||||||||||||
Db   1021 CCCTTCCCGCCAGCGCAG 1038

Qy   961 TGCCCACTGGGCCATAGGCCCTTCTGGGGAGTTCACTGATGCACCTGG 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TGCCCACTGGGCCATAGGCCCTTCTGGGGAGTTCACTGATGCACCTGG 1020

Qy   961 AGTGTGGGCTGAGGCTGGCTCTGAACTGGGACACAGGGGGTCTCAGGAGAGAGCCTTCCTCC 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AGTGTGGGCTGAGGCTGGCTCTGAACTGGGACACAGGGGGTCTCAGGAGAGAGCCTTCCTCC 960

Qy   901 CCCTTCCCAAATGCCTCGGTGTCACTGCACTGCTGTGTGGTAGGGGGTCCCCAACCGGGGCTC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CCCTTCCCAAATGCCTCGGTGTCACTGCACTGCTGTGTGGTAGGGGGTCCCCAACCGGGGCTC 900

Qy   841 AGGGGGGGTGAGGAAGCCCAGCACGTGGGAAGCCTTTTAACCATTCTCGGGGTTGAGCGGAGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AGGGGGGGTGAGGAAGCCCAGCACGTGGGAAGCCTTTTAACCATTCTCGGGGTTGAGCGGAGC 840

; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 5712
; LENGTH: 7086
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5712

Query Match   5.0%;   Score 52.4; DB 10; Length 7086;
Best Local Similarity   69.6%; Pred. No. 2.2e-05;
Matches   71; Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   68 CCCATTTTTCAGATGAGGACTTGAGGCCCCAGAGAGAACGTAAGTAATCTGTCTGAGGCCACA 127
          ||||| || | ||||||| ||||| | ||||| | ||| | | | |||| | ||||||||
Db   4323 CTCATTTTATAGATGAAAACACTGGGGCCTAGAGAAGATAAGTAATCTCTCCAAGGCCACA 4382

Qy   128 CAGCTAGAAAAGCAGCCAGGCCGCCAGAACCCCTGGTCGTGTG 169
          ||||||||||||||||||||||||||||||||||||||||||||
Db   4383 CAGCTAGGAAGTAGCCAGAGCTTATGCCACGTCAGGGGTTTGTGTG 4424

RESULT 5
US-09-764-891-5713
; Sequence 5713, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 5713
; LENGTH: 7388
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5713

Query Match   5.0%;   Score 52.4; DB 10; Length 7388;
Best Local Similarity   69.6%; Pred. No. 2.3e-05;
Matches   71; Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   68 CCCATTTTTCAGATGAGGACTTGAGGCCCCAGAGAGAACGTAAGTAATCTGTCTGAGGCCACA 127
          ||||| || | ||||||| ||||| | ||||| | ||| | | | |||| | ||||||||
Db   4327 CTCATTTTATAGATGAAAACACTGGGGCCTAGAGAAGATAAGTAATCTCTCCAAGGCCACA 4386

Qy   128 CAGCTAGAAAAGCAGCCAGGCCGCCGAACCCCTGGTCGTGTG 169
          ||||||||||||||||||||||||||||||||||||||||||
Db   4387 CAGCTAGGAAGTAGCCAGAGCTTATGCCACGTCAGGGGTTTGTGTG 4428

RESULT 6
US-10-074-024-572/c
; Sequence 572, Application US/10074024
; Publication No. US20030232975A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC001C1
; CURRENT APPLICATION NUMBER: US/10/074,024
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 879
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 572
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-074-024-572

RESULT 7
US-10-027-632-32570
Sequence 32570, Application US/10027632
Publication No. US20020193371A1

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32570
; LENGTH: 710
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(710)
; OTHER INFORMATION: n = A,T,C or G

US-10-027-632-32570

Query Match   4.8%;   Score 49.6;   DB 17;   Length 532;
Best Local Similarity   65.2%;   Pred. No. 0.00011;
Matches 73;   Conservative   0;   Mismatches 39;   Indels   0;   Gaps   0;

Qy   67 TCCCATTTTTCAGATGAGGAGGAGTTGAGGCCCAGAGAACGTAAGTAATCTGTCTGACGCCAC 126
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 TGCCCTTTTAGAGATGAAGGAGACTGAGGCTCAGAGAAGCCAAGTAATTCACCTAAGGCCAC 138

Qy   127 ACAGCTAGAAAGCAGCCAGGCCCAGGCCGGAACCCTGTGTGTCTGCAGCGCCCCA 178
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 GTAGCTAGCAAGAAGCGAGCCAGGACTCAACCTAAGTCTGAGCTCCTGCCA 86

RESULT 8
US-10-027-632-32570
Sequence 32570, Application US/10027632
Publication No. US20030204075A9

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

Query Match   4.8%;   Score 49.6;   DB 13;   Length 710;
Best Local Similarity   72.2%;   Pred. No. 0.00011;
Matches 29;   Conservative   0;   Mismatches 29;   Indels   1;   Gaps   1;

Qy   41 CAGATGTGGTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGA-GTTGAGGCCCAGA 99
          ||||||||||||||| ||| || |||||||||||||| |||||
Db   191 CAGCCTTGTGAGGGGGTCATTTTTATTCCCATTTTACAGATGAAGACACAGAGGCTCAGA 250

Qy   100 GAACGTAAGTAAATCTGTCTGAGGGCCACAGCTAGAAAAGCAGCAGCC 147
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   251 GAGGTCAAGTGATTTGCCCAGGGGCCACACAGCTAGGAGAAACTGCCAGGC 298

RESULT 9
US-10-074-024-71/c
Sequence 71, Application US/10074024
Publication No. US2003023975A1

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC001C1
; CURRENT APPLICATION NUMBER: US/10/074,024
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 879
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (86)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (388)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (391)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (411)
; OTHER INFORMATION: n equals a,t,g, or c

Query Match   4.8%;   Score 49.6;   DB 17;   Length 710;
Best Local Similarity   72.2%;   Pred. No. 0.00011;
Matches 29;   Conservative   0;   Mismatches 29;   Indels   1;   Gaps   1;

Qy   41 CAGATGTGGTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGA-GTTGAGGCCCAGA 99
          ||||||||||||||| ||| || ||||||||||||| |||||
Db   191 CAGCCTTGTGAGGGGGTCATTTTTATTCCCATTTTACAGATGAAGACACAGAGGCTCAGA 250

Qy   100 GAACGTAAGTAAATCTGTCTGAGGGCCACAGCTAGAAAAGCAGCAGCC 147
          ||||||||||||||||||||||||||||||||||||||||||||
Db   251 GAGGTCAAGTGATTTGCCCAGGGGCCACACAGCTAGGAGAAACTGCCAGGC 298

US-10-074-024-71

```
Qy    Query Match
      Best Local Similarity    64.3%;    Score 49.2;    DB 17;    Length 414;
      Matches  72;  Conservative  1;  Mismatches  39;  Indels  0;  Gaps  0;

Qy         67 TCCCATTTTCAGATGAGGAGAGTTGAGGGCCCAGAGACGTAAGTAATCTGTCTGAGGCCAC 126
              ||||| |||||| |||||||||||||||||||||||||||||| |||  | ||||||||||
Db        205 TGCCCTTTTAGAGATGAAGGAGCTCAGAGAAGGCCAAGTAAGTAATTCACTTAAGGCCAC 146

Qy        127 ACAGGTTAGAAGAAAGCAGCCAGCCCGGACCCCTCGGTGTGTCGCAGCCCCCA 178
              ||||||    ||||||| |||||  ||||||||||||| ||||| ||||||||
Db        145 GTAGCTAGCAAGAAAGCGGGGAGCCAGGACTCAACCTAAGTCTGAGCTCCTGCCA 94
```

RESULT 10
US-09-740-033-3/c
; Sequence 3, Application US/0974-0033
; Patent No. US2002010067A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001056
; CURRENT APPLICATION NUMBER: US/09/740,033
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7353
; TYPE: DNA
; ORGANISM: Human
US-09-740-033-3

```
Qy    Query Match
      Best Local Similarity    59.2%;    Score 49.2;    DB 9;    Length 7353;
      Matches  84;  Conservative  0;  Mismatches  58;  Indels  0;  Gaps  0;

Qy         44 ATGTGTTAGGAGGGGCATTAATCAGTCCCATTTTTTCAGATGAGGAGAGTTGAGGGCCCAGAGAAC 103
              ||||| |  ||||  |||||||||||||||  || |||||||||||||||||||||||||||||||
Db       2225 ATGAAGTAGAATAATATTTTTATCTCCATTTTTGCAGAGATGAAGGAGTTCAGAAGG 2166

Qy        104 GTAAGTAATCTGTCTGAGGCCACACAGCTAGAAAGAACAGCCAGCCCGGAACCCCTGG 163
              || |||||||   |   |||  ||  ||||| || ||  ||||| ||  |||| ||||
Db       2165 TTAAGTAACTTACCCAAGGTGCGACAGCTAGTAGGCAAGGGCAGCCACCTACCTGG 2106

Qy        164 TGTGTGCAGCCCCCAGCCCCAGT 185
              |||| |||||||| |||||||||
Db       2105 GCAGTCTGGCGCCCCCAGGTCCAGT 2084
```

RESULT 11
US-09-764-891-7809/c
; Sequence 7809, Application US/0976-4891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7809
; LENGTH: 29163
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7809

```
Qy    Query Match
      Best Local Similarity    67.3%;    Score 48.2;    DB 10;    Length 29163;
```

Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy         51 AGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGGAGAGTTGAGGGCCCAGAGAACGTAAGTA 110
              ||||| ||||||| ||||||| |||||||||||||||||||||||||||||||||  |||||  |
Db       8883 AGGTGAGTGTTGTTATTCCCATTTTACAAAATGAGGAAATTGAGGCTCAGAGAGACGTGAAATT 8824

Qy        111 ATCTGTCTGAGGCCACACACAGCTAGAAAGAACAGCCAGCCCAG 151
              |||||  | ||||  || ||||||   | ||  |||||||||||
Db       8823 ACCAGTCCAAGATTACACAGCTTAATGAGGCTGCAGAGTCCAG 8783
```

RESULT 12
US-10-741-600-34521
; Sequence 34521, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34521
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-34521

```
Qy    Query Match
      Best Local Similarity    58.7%;    Score 48;    DB 21;    Length 201;
      Matches  81;  Conservative  1;  Mismatches  56;  Indels  0;  Gaps  0;

Qy         48 GGTAGGAGGGGCATATTCAGTCCCATTTTTCAGATGAGGAGAGTTGAGGGCCCAGAGAACGTAA 107
              |||| ||| |||| |||||||||||||| || ||||||||| |||||| |||||||| |||||
Db          8 GGAGGTAGGAACTATATATCCCCATTTTCTTAGATGAAGGAAAACTGAGGCTCAGGGAGGAA 67

Qy        108 GTAATCTGTCTGAGGCCACACAGCTAGAAAGAACAGCCAGCCCGGAACCCCTGGTGTG 167
              || |||| |   | |  ||  ||||| || || || |||||||  ||||| || |||
Db         68 GAAATTTACCCATGGTCACAAAAGCTAGTAGTGAGTCAAATCCAAAACCAGGTATG 127

Qy        168 TGCAGCCCCCAGCCCCAGT 185
              ||||||||||  ||||||
Db        128 CGTAACACCAAAAGCCAGT 145
```

RESULT 13
US-10-027-632-226737/c
; Sequence 226737, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
```

RESULT 14
US-10-027-632-226738; Application US/10027632
Sequence 226738, Application US/10027632
Publication No. US20020198371A1

GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226738
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226738

Query Match         58.7%;   Score 48;  DB 13;  Length 585;
Best Local Similarity   4.6%;  Pred. No. 0.00034;
Matches  81;  Conservative  1;  Mismatches  56;  Indels  0;  Gaps  0;

Qy    48 GGTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCCAGAGAACGTAA 107
         ||||  |||| || ||||||||||| |||||||  |||||| |||| | ||| ||||||||
Db   493 GGAGGTAGGAACTATTATCCCCATTTTTCTAGATGAAGAAACTGAGGCTCAGGGAAGGAAA 434

Qy   108 GTAATCTGTCTGAGGCCACACAGCTAGAAAAGCAGCCAGGCCGACCCCTTGGTGTG 167
         |||| ||| | |    ||| ||||||| ||||| | | | ||| |||| | ||| |
Db   433 GAAATTTACCCATGGTCACAAAGCTAGTGAGTASTGAGCTCAAATCAAAACCCAGGTATG 374

Qy   168 TGCAGCCCCCAGCCCAGT 185
         ||| |  |||| ||||||
Db   373 CGTAACACCAAAGCCAGT 356

RESULT 15
US-10-027-632-226739/c
Sequence 226739, Application US/10027632
Publication No. US20020198371A1

GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226739
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226739

Query Match         58.7%;   Score 48;  DB 13;  Length 585;
Best Local Similarity   4.6%;  Pred. No. 0.00034;
Matches  81;  Conservative  1;  Mismatches  56;  Indels  0;  Gaps  0;

Qy    48 GGTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCCAGAGAACGTAA 107
         ||||  |||| || ||||||||||| |||||||  |||||| |||| | ||| ||||||||
Db   493 GGAGGTAGGAACTATTATCCCCATTTTTCTAGATGAAGAAACTGAGGCTCAGGGAAGGAAA 434

Qy   108 GTAATCTGTCTGAGGCCACACAGCTAGAAAAGCAGCCAGGCCGACCCCTTGGTGTG 167
         |||| ||| | |    ||| ||||||| ||||| | | | ||| |||| | ||| |
Db   433 GAAATTTACCCATGGTCACAAAGCTAGTGAGTASTGAGCTCAAATCAAAACCCAGGTATG 374

Qy   168 TGCAGCCCCCAGCCCAGT 185
         ||| |  |||| ||||||
Db   373 CGTAACACCAAAGCCAGT 356

Search completed: June 29, 2005, 18:26:05
Job time : 762 secs

OM nucleic - nucleic search, using sw model

```
                          GenCore version 5.1.6
                          Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:                   June 29, 2005, 15:10:02 ; Search time 4078 Seconds
                                                    (without alignments)
                                                    9688.747 Million cell updates/sec

Title:                    US-10-617-443B-1_COPY_1_1038
Perfect score:            1038
Sequence:                 1 cagggactcgaagttcgcag........ggcccttcccgcgcagcgcag 1038
Scoring table:            IDENTITY NUC
                          Gapop 10.0, Gapext 1.0
Searched:                 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters:    68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :                EST:*
                          1: gb_est1:*
                          2: gb_est2:*
                          3: gb_htc:*
                          4: gb_est3:*
                          5: gb_est4:*
                          6: gb_est5:*
                          7: gb_est6:*
                          8: gb_gss1:*
                          9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Match Length	DB	ID	Description
c	1	55.8	5.4	421	1	AL119813	AL119813 DKFZp761H
c	2	55.8	5.4	458	1	AL119782	AL119782 DKFZp761F
c	3	54.8	5.3	925	9	CNS00091P	CNS00091P Drosophil
c	4	51.6	5.0	547	8	AQ808226	AQ808226 HS_4554_A
	5	50.6	4.9	383	2	BE099553	BE099553 UI-R-BJ1-
	6	50.6	4.9	449	1	CC773356	CC773356 CH240_17D
	7	50	4.8	536	9	CE070706	CE070706 tigr-gss-
	8	49.6	4.8	689	9	AG113318	AG113318 Pan trogl
c	9	48.6	4.7	790	6	CC552420	CC552420 CH240_438
c	10	48	4.6	667	8	CB321964	CB321964 CH240_438
c	11	46.8	4.5	454	8	AQ415695	AQ415695 RPCI-11-1
c	12	46.8	4.5	672	8	AQ237490	AQ237490 RPCI11-69
c	13	46.8	4.5	925	9	CNS00091P	CNS00091P Drosophil
	14	46.6	4.5	150	2	BE145776	BE145776 IL5-HT020
	15	46.6	4.5	422	4	BI028582	BI028582 IL5-MT020
	16	46.6	4.5	510	9	AQ479326	AQ479326 RPCI-11-2
c	17	46.4	4.5	700	8	CB466727	CB466727 tigr-gss-
c	18	46.2	4.5	643	7	CN070055	CN070055 857546 MA
c	19	45.6	4.4	586	2	AW658542	AW658542 94632 MAR
c	20	45.4	4.4	296	8	AA296443	AA296443 EST11220
c	21	45.2	4.4	662	9	AG051198	AG051198 Pan trogl
	22	45	4.3	280	1	AI702247	AI702247 we04c04.x
	23	45	4.3	462	1	AI884801	AI884801 wl84h03.x
	24	44.8	4.3	506	8	AQ777662	AQ777662 HS_2252_B
c	25	44.6	4.3	460	2	BF921020	BF921020 MR2-NT013
c	26	44.6	4.3	575	5	BQ364741	BQ364741 MR2-SN000
c	27	44.6	4.3	578	2	BE109848	BE109848 UI-R-CA0-
c	28	44.6	4.3	910	9	BG378055	BG378055 UI-R-CT0-
c	29	44.6	4.3	668	1	AV221826	AV221826 AV221826
c	30	44.6	4.3	733	9	AG030822	AG030822 Pan trogl
	31	44.6	4.3	762	9	CD777006	CD777006 UI-M-AQ0-
	32	44.6	4.3	763	9	CD774640	CD774640 UI-M-AQ0-
c	33	44.4	4.3	253	1	AI560342	AI560342 tn13c04.x
c	34	44.4	4.3	591	1	AU145409	AU145409 AU145409
c	35	44.4	4.3	750	9	CC531733	CC531733 CH240_408
c	36	44.2	4.3	480	9	CE347071	CE347071 tigr-gss-
	37	44.2	4.3	910	9	AL065629	AL065629 Drosophil
c	38	44.2	4.2	610	8	AQ374046	AQ374046 RPCI11-15
c	39	44	4.2	717	9	CE168278	CE168278 tigr-gss-
	40	43.8	4.2	261	2	AW270824	AW270824 xp54h06.x
	41	43.8	4.2	742	5	BP159679	BP159679 BP159679
c	42	43.8	4.2	769	6	CD104579	CD104579 AGENCOURT
c	43	43.8	4.2	836	9	CC485946	CC485946 CH240_315
c	44	43.8	4.2	438	8	AQ010437	AQ010437 HS_2167_B
	45	43.6	4.2	504	8	B94987	B94987 CIT-HSP-216

ALIGNMENTS

```
RESULT 1
AL119813/c
LOCUS       AL119813
DEFINITION  DKFZp761H1124_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
            DKFZp761H1124 5', mRNA sequence.
ACCESSION   AL119813
VERSION     AL119813.1  GI:5925712
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Bloecker, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp761H1124) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rpd.de.
FEATURES             Location/Qualifiers
     source          1..421
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761H1124"
                     /tissue_type="amygdala"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="761 (synonym: hamy2)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN
```

```
            Query Match     5.4%;    Score 55.8; DB 1; Length 421;
            Best Local Similarity   70.1%;    Pred. No. 0.0027;
            Matches  89; Conservative  0; Mismatches  37; Indels  1; Gaps  1;
```

RESULT 2

```
Qy   67 TCCCATTTTTCAGATGAGGAGTTGAGGCGCCCAGAGAACGTAAGTAATCTGTCTTGAGGCCAC 126
        ||||||||| | |||||||||  ||||||| | |||||  ||   || ||||| |||||||||
Db  279 TCCCCATTTTCAGGTGTTGAAACTGAGGCGCCAGAGACGTTTAGTGACTTGCTCAAGGTCAC 220

Qy  127 ACAGCTAGAAAACCAGCCAGGCC-CAGCCGAACCCCTGGTGTGTGTGTCAGCCCCCAGT 185
        ||||||| |  || ||  ||||||  ||||  || ||  | |||  | |||||| | ||
Db  219 ACAGCTCAAAAGCAGCAAAAGCCCGAAAGCTGAAACCCCAGGTTCGTCTCCCCTTGGCCAGG 160

Qy  186 TGCTCAT 192
        || |||||
Db  159 CCATCTT 153
```

RESULT 2

```
Query Match    5.4%;   Score 55.8;   DB 1;   Length 458;
Best Local Similarity   70.1%;   Pred. No. 0.0027;
Matches  89;   Conservative   0;   Mismatches  37;   Indels   1;   Gaps   1;

Qy   67 TCCCATTTTTCAGATGAGGAGTTGAGGCGCCCAGAGAACGTAAGTAATCTGTCTTGAGGCCAC 126
        ||||||||| | |||||||||  ||||||| | |||||  ||   || ||||| |||||||||
Db  279 TCCCCATTTTCAGGTGTTGAAACTGAGGCGCCAGAGACGTTTAGTGACTTGCTCAAGGTCAC 220

Qy  127 ACAGCTAGAAAACCAGCCAGGCC-CAGCCGAACCCCTGGTGTGTGTGTCAGCCCCCAGT 185
        ||||||| |  || ||  ||||||  ||||  || ||  | |||  | |||||| | ||
Db  219 ACAGCTCAAAAGCAGCAAAAGCCCGAAAGCTGAAACCCCAGGTTCGTCTCCCCTTGGCCAGG 160

Qy  186 TGCTCAT 192
        || |||||
Db  159 CCATCTT 153
```

RESULT 3

```
Query Match    5.3%;   Score 54.8;   DB 9;   Length 925;
Best Local Similarity   16.5%;   Pred. No. 0.0057;
Matches  60;   Conservative  161;   Mismatches 142;   Indels   1;   Gaps   1;

Qy  488 GGGCCCAGAGGGGCCCACGCGTGTGTCCGGGCGCCTGGGTAGTCTCAGAAGGGTCACCTGGCGGTC 547
        ::||::||||||||||:|::|||:|::||||::||::|:||||::|||||||||||||::::|||||
Db  914 SSBSCCSSSSMSTSSSSNSSBCSSCSSSBSSSSTSSMSSSSGSSSSSSSGTSSACVKC 855

Qy  548 TTCCACTACACCCCCGCCTGGACACTGCTGTAGCCCCAGGGCTCGGAGGGACCAGCTGGA 607
        :::|||::|||||||:||:::||||:||::|||::|||||||::||||||||::||||||
Db  854 NASSSCGCGCGWABCCMCSSSSSCCGSASARGVKVRASGGGAGKRGGGSSGGASASHSSSS 795

Qy  608 GCCCATGAGGAGGAGGGGCCCAGTTCTTCTCTGTAAGGGGTATTGCTGTAGCATGAGGGAAACAG 667
        ::|||||||||||||||||||:||||:|:|:|::||||||:|:||||::||:||::|||||||
Db  794 ACBSSSSSCSASCWSASSSSSSASSSRSRGGGAGGSGASSSSSSSSASAGSVVSSAS 735

Qy  668 ACAAGGCCCAGGGGGGGGACTAACCCGAGATCCAGCCCCGGCCTCACTCCCGTGGTCACG 727
        ||:::||||||||::|||::|:||:|:|||||||:||::||||||||:||:|||||||||
Db  734 SSSSSCCSSSVSCSSSVASMSCSSBBSSSSASASSSSSSSSCASCSCSCCCTSWSCCSSTSA 675

Qy  728 GCAATATCCTTAACCTTCTCTCTGAGCCTCCTGCCCAGGGCTTAGACAGGGTCCAGTGAGGGGGG 787
        :::||::::||||::|||:|:|:||||||||:|||||||:||||:||||||::||||||||||||
Db  674 SMSA-ARSSSSSSSSSCSSSSMSASSSASSSSSSSASSSGSSSSSGSSSSSSSACBSMSSGGGSGS 616

Qy  788 TGAGGAAAGCCCCAGCACGGTGGAAAAGCCTTTTTAACCATTCTCGGGGTGAGCGGGAGCGGCTTCC 847
        :||||||||||||||||:||||||||:|::||:::::|::||::|:::|:::|:||||::|||
Db  615 VSASSGMSSSVSSSGGRSSSGGSGGGGGGGGGGGVGSGSGSGSGSGGCSCSSSGCMRCSCS 556
```

RESULT 4

Qy 848 CAAA 851
        ||||
Db 555 SAAA 552

LOCUS   AQ808226
DEFINITION HS_4554_A2_H04_T7A CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=4554 Col=8 Row=O, genomic survey
            sequence.
ACCESSION   AQ808226
VERSION   AQ808226.1 GI:5727468
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 547)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE   Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 4554 row: O column: 8
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 547.
FEATURES   Location/Qualifiers
     source   1..547
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=4554 Col=8 Row=O"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Query Match   5.0%; Score 51.6; DB 8; Length 547;
Best Local Similarity   60.9%; Pred. No. 0.034;
Matches   54; Conservative   0; Mismatches   54; Indels   0; Gaps   0;

Qy   48 GGTRAGGAGGGCATRATTCAGTCCCATTTTTTCAGTATGAGGGCCCAGAGAACGTAA 107
         ||| || || |||| ||||| || || || ||||| | | ||||| | |||
Db 225 GGAGGTAGGAACTATTATTCCCCATTTCTRGATGAAGAAACTGAGGGCTCAGGGAAGGAA 284

Qy 108 GTRATCTGTCGTCTGAGGGCCACAGCTAGRAAAGCACAGCCCGGAACCCTGGTGTG 167
         ||||| ||| |||||| | | |||||||| | | || || || |||||||| ||
Db 285 GAAATTACCCATGGTCACAAAGCTAGTAGTGAGCTCAAAATCCAAAACCCAGGTATG 344

Qy 168 TGCAGCCCCCAGCCCCAGT 185
         |||| || ||| ||||||
Db 345 CGTAACACCAAAAGCCAGT 362

RESULT 5

LOCUS   BE099553
DEFINITION UI-R-BJ1-atj-f-11-0-UI.s1.UI-R-BJ1 Rattus norvegicus cDNA clone
                              383 bp   mRNA   linear   EST 13-JUN-2000
ACCESSION   BE099553
VERSION   BE099553.1 GI:8491431
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 383)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized ventricle at 16.5 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..383
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ1-atj-f-11-0-UI"
                     /clone_lib="UI-R-BJ1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
                     library is a subtracted library derived from the UI-R-BJ1
                     library. The subtracted library derived from the following
                     tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
                     canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
                     AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
                     For a detailed description of the library from which this
                     clone was derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_TISSUE=ventricle at 16.5 dpc
                     TAG_LIB=UI-R-BJ1
                     TAG_SEQ=GTTCG"
ORIGIN

Query Match   4.9%; Score 50.6; DB 2; Length 383;
Best Local Similarity   70.1%; Pred. No. 0.058;
Matches   68; Conservative   0; Mismatches   29; Indels   0; Gaps   0;

Qy   51 AGGAGGGCATATTCAGTCCCCATTTTTTCAGATGAGAGGAGTTGAGGGCCCAGAGAACGTAAGTA 110
         ||||||| |||||| ||| ||||| | |||| | | | || ||| | ||| | | |
Db   23 AGGTGGTCATTTTTATTCCCATTGTGTCTAGTTGAGRAAACAGAGGGCTCAGAGGAAGCCAAGTG 82

Qy 111 ATCTGTCTGAGGGCCACACAGCTAGAAGCAGCCAGGC 147
         ||||||| |||||||| ||||||||||||| |||||||
Db   83 AGTCGCCCAGAGCCACACACACGCTAGAAAAGTTGCCAGGC 119

RESULT 6

LOCUS   CC773356

```
LOCUS       CC773356                449 bp    DNA     linear   GSS 27-JUN-2003
DEFINITION  CH240_17D11.TV CHORI-240 Bos taurus genomic clone CH240_17D11,
            genomic survey sequence.
ACCESSION   CC773356
VERSION     CC773356.1  GI:32325647
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
            Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
            Womack,J.E., de Jong,P.J. and Lewin,H.A.
  TITLE     Bovine BAC end sequences from CHORI-240 library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Harris Lewin
            Department of Animal Sciences
            University of Illinois at Urbana Champaign
            1201 W. Gregory Dr., Urbana, Il 61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5517
            Email: h-lewin@uiuc.edu
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/ordering_information.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by by University of Illinois at Urbana
            Champaign, USA with funds provided by grant No. AG202-34480-11828
            from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
            Genome Sequencing Initiative)
            Plate: 17 row: D column: 11
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_17D11"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                     Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
ORIGIN
```

```
RESULT 7
CE070706
LOCUS       CE070706                536 bp    DNA     linear   GSS 24-SEP-2003
DEFINITION  tigr-gss-dog-1700023012295 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE070706
VERSION     CE070706.1  GI:35132048
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canoidea; Canidae; Canis.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..536
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site 1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN
```

```
Query Match             4.9%;  Score 50.6;  DB 9;  Length 449;
  Best Local Similarity   60.6%;
  Matches   83;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         33 AGTTCCCACAGATGTGGTPAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAG 92
              |||| | |||||||||| ||||||||||||||||||||||| |||||||||||||  |||
Db         81 AGCTTGCACCGACCTCGTTGAGAAACGTPAATGTTAGCCCATTTTTACAGATGAGGAACTGAG 140

Qy         93 GCCCAGAGAACCTPAAGTPAATCTCTGTCTGAGGGCCACACAGCTCTGAAAAAGCACCCAGC 152
              |||||||||  ||||| |||| | |||| ||||| |||||||||| |||| |||||||||
Db        141 GCTCAGAGAGGTTGAGTCATTGAGCTGAGGTCACACACTAGCACTGCGTCTGGGG 200

Qy        153 CGAACCCCTGGTGTGTG 169
              | ||||||| ||||| |
Db        201 AGGAGCCCGAGTCGCTG 217
```

```
RESULT 8
AG113318
LOCUS       AG113318                689 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-119015.R, genomic survey sequence.
ACCESSION   AG113318
VERSION     AG113318.1  GI:16733837
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 689)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
```

```
Query Match             4.8%;  Score 50;  DB 9;  Length 536;
  Best Local Similarity   60.1%;
  Matches   83;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         49 GTAGGAGGGCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGGCCCAGAGAACGTAAG 108
              | ||| || || ||||| ||| || ||   |||||| ||   ||||| |||| ||| |||
Db         36 GCAGAAGGAGATATTTTGAGCGTGATCTTACAAAAAGAGGAGAAACGTGAGCTCAGAGACCT 95

Qy        109 TAATCTGTCTGAGGGCCACACAGCTAGAAAGCACGCCAGCCGGAACCCCTGGGTGTGT 168
              || | | ||| || ||||||||| || ||| || |||  ||  ||||  || |||||
Db         96 TAGCTAGCCCAAGGTCACACACAGCTAGTAGTAAGTGGCCCAGCTGAGATTCAACTTTGAGATT 155

Qy        169 GCAGCCCCCAGCCCAGTT 186
              |||| || ||| | |
Db        156 GAGGCCCCAAATCCGTGCT 173
```

COMMENT
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

FEATURES
  source   1..689
     /organism="Pan troglodytes"
     /mol_type="genomic DNA"
     /db_xref="taxon:9598"
     /clone="PTB-119015.R"
     /sex="male"
     /cell_type="lymphoblast"
     /clone_lib="PTB Chimpanzee Male BAC Library"

PRIMERS
  Sequencing: M13Rev
LIBRARY
  Vector : pKS145
  R.Site 1 : SacI
  R.Site 2 : SacI

ORIGIN

RESULT 9
CC552420/c
LOCUS   CC552420   790 bp   DNA   linear   GSS 17-JUN-2003
DEFINITION   CH240_438G3.TARBAC13P2 CHORI-240 Bos taurus genomic clone
CH240_438G3, genomic survey sequence.
ACCESSION   CC552420
VERSION   CC552420.1   GI:31870704
KEYWORDS   GSS.
SOURCE   Bos taurus (cow)
  ORGANISM   Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE   1 (bases 1 to 790)
AUTHORS   Holt,R., Stott,J., Yang,G., Barber,S., Smailus,D., Prabhu,A.-L.,
Teai,M., Cloutier,A., Lee,D., Girn,N., Olson,T., Mayo,M.,
Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R.,
Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S.,
Schein,J., Marra,M., de Jong,P., Keele,J.W. and Kappes,S.M.
TITLE   Bovine BAC End Sequences from library CHORI-240, PLATES 399 to 478
JOURNAL   Unpublished (2003)
COMMENT   Contact: Rob Holt
Other_GSSs: CH240_438G3.T7
Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6005
Fax: 604-877-6276
Email: rholt@bcgsc.ca
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering_information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the

Query Match   4.8%; Score 49.6; DB 9; Length 689;
Best Local Similarity   72.2%; Pred. No. 0.12;
Matches 78; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

Qy   41 CAGATGTGTGGTAGGAGGGCCATATTCAGTCCCATTTTTCAGATGAGGA-GTTGAGGGCCAGA 99
Db   185 CAGCCTTGTGTGAGGGGGTCATTTTTATTCCCATTTTTACAGATGAGAACACAGAGGGCTCAGA 244
Qy   100 GAACGTAAGTAGAATCTGTCTGAGGGCCACACACAGCCTAGAGAAAGCAGGCCAGGC 147
Db   245 GAGGTCAAGTCAGTGATTTGCCCAGGGGGCCACACAGCCTAGGAAAACTGCCAGGC 292

FEATURES
  source   1..790
     /organism="Bos taurus"
     /mol_type="genomic DNA"
     /strain="breed: Hereford"
     /db_xref="taxon:9913"
     /sex="Male"
     /clone="CH240_438G3"
     /cell_type="Blood"
     /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull Li Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

British Columbia Genome Sciences Centre, Canada.
Plate: 438 row: G column: 3
Seq primer: SP6
Class: BAC ends.

ORIGIN

RESULT 10
CB321964
LOCUS   CB321964   667 bp   mRNA   linear   EST 05-MAR-2003
DEFINITION   UI-CF-FN0-aem-m-06-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone
UI-CF-FN0-aem-m-06-0-UI 3', mRNA sequence.
ACCESSION   CB321964
VERSION   CB321964.1   GI:28856622
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 667)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED   8889548
COMMENT   Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 174-405, >MIR#SINE/MIR (matched compliment)
Seq primer: M13 FORWARD
PolyA=Yes.

Query Match   4.7%; Score 48.6; DB 9; Length 790;
Best Local Similarity   55.7%; Pred. No. 0.22;
Matches 93; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy   48 GGTAGGAGGGCCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGGCCCAGAGAACGTAA 107
Db   646 GGGAGGTAGGTATTCTTAATCCCATTTTGTCAGATGAGTAATTGAAACTCAGAGAGGTTAA 587
Qy   108 GTAATCTGTCTGAGGGCCACACAGCCTAGAAAAGCAGCCAGGCCGGAACCCCTGGTGTG 167
Db   586 GTCACTTGCCCATAGTCACACAGCTGGTTAATGACAAATTCTGGCTGCAAACCCAGCTGT 527
Qy   168 TGCAGCCCCAGCCCAGTTGCTCATTGCGGGGGCTCGGGGAGCCACGAG 214
Db   526 TACTGACTCCCCATCTGAGGTTTTTTCCATTATGGTGTGTGCTACCAG 480

FEATURES   Location/Qualifiers
     source   1..667
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clones="UI-CF-FN0-aem-06-0-UI"
                     /tissue_type="Human Lung Epithelial cells"
                     /lab_host="DH10B (Life Technologies) (Tn phage resistant)"
                     /clone_lib="UI-CF-FN0"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     /note"Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     DUI) The library was subtracted according to according to
                     normalized Human lung epithelial cell libraries (EN1 and
                     UI-CF-FN0 is a subtracted cDNA library derived from two
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. For additional information, contact:
                     bento-soares@uiowa.edu
                     TAG TISSU=Lung Epithelial Cells Tissue nos 359-368
                     TAG_LIB=UI-CF-FN0
                     TAG_LIB=UI-CF-FN0
                     TAG_SEQ=GGCTGTAGGC"
ORIGIN
     Query Match   4.6%;   Score 48;   DB 6;   Length 667;
     Best Local Similarity   59.6%;   Pred. No. 0.3;
     Matches   81;   Conservative   0;   Mismatches   55;   Indels   0;   Gaps   0;

Qy   51 AGGAGGGGCATATTCAGTCCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGGAACGTTAGTA 110
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   264 AGGATGGGAGAATCAGCCCCACTTTTGAGATGATCTTGGAAGGCTCAGAGAGGAGGTACAGA 323

Qy   111 ATCTGTCTCTGAGGCCACACAGCTAGAAAAGCAGCCAGGCCCAAACCCCTGGTGTGTGTGC 170
              ||||||||| ||||||||| |||||||||||||| |||| ||||||||||||||
Db   324 AGTGTGTCTGAGGACTCGCAGCGGGGAAGGATGCCCTGAGCCATTTCAAAACCAGGCAGTCC 383

Qy   171 AGCCCCCAGCCCCAGTT 186
              ||||||||||||||||
Db   384 TGCATACAAGTCCGTT 399

RESULT 11
LOCUS   AQ415695   454 bp   DNA   linear   GSS 23-MAR-1999
DEFINITION   RPCI-11-196N2.TV RPCI-11 Homo sapiens genomic clone RPCI-11-196N2,
             genomic survey sequence.
ACCESSION   AQ415695
VERSION   AQ415695.1 GI:4474664
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 454)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
             Venter,J.C.
  TITLE   Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
             Map Building
  JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: RPCI-11-196N2.TJ
             Contact: Shaying Zhao, William Nierman, Mark Adams
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: hbe@tigr.org
             Clones are derived from the human BAC library RPCI-11. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
             Research Genetics (info@resgen.com). BAC end search page:
             http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
             Seq primer: T7
             Class: BAC ends.

FEATURES   Location/Qualifiers
     source   1..454
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7575193"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-196N2"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC library"
ORIGIN
     Query Match   4.5%;   Score 46.8;   DB 8;   Length 454;
     Best Local Similarity   58.7%;   Pred. No. 0.57;
     Matches   81;   Conservative   0;   Mismatches   57;   Indels   0;   Gaps   0;

Qy   48 GGTAGGAGGGGCATATTCAGTCCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGGAACGTAA 107
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   165 GGAGGTGAGGGAACTATTATCCCCATTTTCTAGATGAAGAAACTGAGGGCTCAGGGGAAGGAAA 224

Qy   108 GTAATCTGTCTCTGAGGCCCACACAGCCTTAGAAAAGCCAGGCCCAGGCGGAACCCTGGTGTG 167
              ||||| |||||||||||||| ||||| ||||||||||||||||||| || |||||||
Db   225 GAAAATTTACCCATGGGTCACAAAGCTAGTGAGTGAGCTCAAAATCCAAACCCAGGTATG 284

Qy   168 TGCAGCCCCAGCCCAGT 185
              |||||||||||||||
Db   285 CGTTACACCAAAGCCAGT 302

RESULT 12
LOCUS   AQ237490   672 bp   DNA   linear   GSS 21-APR-1999
DEFINITION   RPCI11-69K22.TK RPCI-11 Homo sapiens genomic clone RPCI-11-69K22,
             genomic survey sequence.
ACCESSION   AQ237490
VERSION   AQ237490.1 GI:3669781
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 672)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
             Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE   Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: RPCI11-69K22.TJ
             Contact: Mark Adams
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: mdadams@tigr.org
             Clones are derived from the human BAC library RPCI-11. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
             Research Genetics (info@resgen.com). BAC end search page:
             http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
             Seq primer: T7
             Class: BAC ends.

FEATURES   Location/Qualifiers
     source   1..672
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7526373"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-69K22"
                     /sex="Male"
                     /cell_type="Lymphocytes"

RESULT 13

LOCUS   CNS0091P   925 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC #
             BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION   AL053013
VERSION   AL053013.1 GI:4934461
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 925)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see http://www.fruitfly.org The BDGP Drosophila
             melanogaster BAC library was prepared by Kazutoyo Osoegawa and
             Aaron Mammoser in Pieter de Jong's laboratory in the Department of
             Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
             NY. The library is named RPCI-98 and was constructed by partial
             EcoRI digestion of Drosophila DNA provided by the BDGP from the
             isogenic strain y2; cn bw sp, the same strain used for the BDGP's
             P1 and EST libraries. A more detailed description of the library
             and how to order individual BAC clones, the entire library, or
             filters for hybridization from the BACPAC Resource Center can be
             found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES   Location/Qualifiers
     source   1..925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="TET3"
ORIGIN
Query Match   4.5%; Score 46.8; DB 9; Length 925;
Best Local Similarity   13.0%; Pred. No. 0.66;
Matches   48; Conservative 164; Mismatches 156; Indels 0; Gaps 0;

Qy   138 GCAGCCAGGCCGCCAGCCGAACCCCTGGGTGTGGTGCAGCCCCCCAGCCCAGCCCAGTTGCTCATTGCGG 197
          :: :::|||:|:::|| :||||::||:::| |:|:|||||||| ::||||||| ||::|| :::
Db   555 SSGSGYGKGCSGSGGBSSCCCSSCSCSCCBCCCCSCSCSSYCCSSSBSSSSKCSSTS 614

Qy   198 GGCTCGGGACCACCAGGACGCGAGGCGTGAGCAGCCATGTGTTCCAGATCGGTCGGAACTCGAGAG 262
          ::|| : ::: :::   : : ::: : :||:: |:: :: :: | :: :
Db   198 GAGATCTTAAGTAATTTACCTAAGGTCACACAACTAATTAATTAATTGGCAGGGCGAG 315

RESULT 14

LOCUS   BE145776   150 bp   mRNA   linear   EST 21-JUN-2000
DEFINITION   IL5-HT0207-291099-007-A11 HT0207 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE145776
VERSION   BE145776.1 GI:8608500
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 150)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
             O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2704922
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=IL5-HT0207-291
             099-007-A11&t3=1999-10-29&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 150.
FEATURES   Location/Qualifiers
     source   1..150
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HT0207"
                     /clone_lib="HT0207"
                     /note="Organ: head.neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under

ORIGIN
Query Match   4.5%; Score 46.8; DB 8; Length 672;
Best Local Similarity   63.2%; Pred. No. 0.62;
Matches   72; Conservative   0; Mismatches   42; Indels   0; Gaps   0;

Qy   38 CCACAGATGTGGTAGGAGGGGGGCATATTCAGTCCCATTTTTTTCAGATGAGGAGGTTGAGCCCCA 97
          ||:|||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   615 BSCSCCCSSKSVCGTCSCSSSSSSCCSSSSSSTSSSSSSTSSSTSSTSKSSSSSSSSSSSSYTTSKS 674

Qy   98 GAGAACGTAAGTAATCTCTCTGAGCCCCCACACAGCTAGAAAGCAGCCAGGCCCAG 151
          ||||| ||||||:::||:||: :||:||||||||||||:||||||:||||||:||||
Db   258 AGCCCGGCCACAGGCCCGTGCAGGGAACCCCGGAGGGCTGTAGGCCCCGTGCCAC TGCATGC 317

Qy   202 CAACAACTCTATGAGATAAAAACAATTATTTCCATTTACAGTCAAGAGAATTGAGCGCACA 261
          |||||||||:||:||||||||||||:||:|||||||||||||||||||||||||||||||||
Db   675 TSASGSGSWSWSAGGGSGSTGSTSSSSSSSTSTSSSSVSSGSKSSTBSSSGSBSSSGSGSSSS 734

Qy   262 GAGATCTTAAGTAATTTACCTAAGGTCACACAACTAATTAATTAATTGGCAGGGCGAG 315

Db   318 CTCAGGGCCGTTGGGTGTCTTGGCAGCCCACACAGCCCCCTACTGTCTGACGGCAGCCAGCAGCAGGAAATCTGAGC 377
          |:::|||||:||||| |:|||||||||:|||||:||||||||||::||||::||:||| ||||||||::||| :||
Qy   378 CCGGGAAGGGTTCAGGGAAGAGTTCGTGAACCATCTAGCCAAGTCGGGCTGGGGTGGGGCCAA 437

Db   735 STSSBBBCTSTSTSSSSSSYSSSSSTCSSCCTCCSYSYSSSSTSSSSTSSWGSTSGSSSSSVGT 794

Qy   438 GTTAGACACAGATGTAGGGGCCCTGTGGACTCAGAAATTGGCAGCTCTTTTGGCCCCAGAGG 497
          ||||:|:||||||||||||| :||||||||||:| |::|||| :::||||||||:|| :|
Db   795 SSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCCGGSSSGKGGVVTKCGCGCGCGSSSTN 854

Qy   498 GGCCACGC 505
          |:::|:
Db   855 GMBGTSSACSSSSSSCSSSSVSSSSRKSSASSSSSVSSSGSSGGVSNSNSSSASRKSSSSSGSVSS 914

Qy   915 GSGSGSGS 922

```
ORIGIN                    low stringency conditions."

Query Match            4.5%;    Score 46.6;  DB 2;  Length 150;
Best Local Similarity  71.8%;   Pred. No. 0.51;
Matches    61;  Conservative     0;  Mismatches   24;  Indels    0;  Gaps    0;
.

Qy      51 AGGAGGGCCATATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAGAACGTAAGTA 110
           ||||  |||||||||||||||||||||||||||||||||||| |||||  |||||||||||
Db      49 AGGTAGGGCACTATTAGACCCACTTTCAGATGAGGAAGAAAGGCTCAGGAGAGGTTAACTA 108

Qy     111 ATCTGTCTGAGGGCCACACAGCTAGA 135
           ||||| ||||| ||| ||||||| ||
Db     109 ATTTGTCCAAGGTCACACAGCAGGA 133

Qy      62 TTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAGAACGTAAGTAAGTAATCTGTCTGAG 121
           ||| |||||| ||||||||||||||||||||||||| |||  |||| |||||  |||||  ||
Db     206 TACAATTCCATTTTTCAGATCAGGAAAACCAGGGCTCAGGAGAAGTTAAGTAACTTGCCCAAG 147

Qy     122 GCCACACAGCTAGAAAG 138
           |||||||||||||||||  |||
Db     146 GTCACACAGCCAGTAAG 130

Search completed: June 29, 2005, 18:09:38
Job time : 4087 secs


RESULT 15
BI028582/c
LOCUS        BI028582              422 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION   IL5-MT0208-210201-356-f06 MT0208 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI028582
VERSION      BI028582.1  GI:14435212
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 422)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
             O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
PUBMED       10737800
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=IL5&t2=IL5-MT0208-
             210201-356-f06&t3=2001-02-21&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 320.
FEATURES             Location/Qualifiers
     source          1..422
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0208"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No.
                     196,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
ORIGIN

Query Match            4.5%;    Score 46.6;  DB 4;  Length 422;
Best Local Similarity  75.3%;   Pred. No. 0.64;
Matches    58;  Conservative     0;  Mismatches   19;  Indels    0;  Gaps    0;
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 19:59:10 ; Search time 202.461 Seconds
                   (without alignments)
                   2153.984 Million cell updates/sec

Title: US-10-617-443B-2_COPY_1_9
Perfect score: 44,
Sequence: 1 MHIALPASA 9

Scoring table: BLOSUM62
                   Xgapop 10.0 , Xgapext 0.5
                   Ygapop 10.0 , Ygapext 0.5
                   Fgapop 6.0 , Fgapext 7.0
                   Delop   6.0 , Delext   7.0

Searched: 4708233 seqs, 2422760795 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10617443/runat_29062005_135931_26061/app_query.fasta_1.718
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10617443 @CGN_1_1_3890 @runat_29062005_135931_26061 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -THREADS=1 -XGAPOP=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
                   1: gb_ba:*
                   2: gb_htg:*
                   3: gb_in:*
                   4: gb_om:*
                   5: gb_ov:*
                   6: gb_pat:*
                   7: gb_ph:*
                   8: gb_pl:*
                   9: gb_pr:*
                  10: gb_ro:*
                  11: gb_sts:*
                  12: gb_sy:*
                  13: gb_un:*
                  14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	% DB	ID	Description
1	44	100.0	60597	9	AC113611	AC113611 Homo sapi
2	39	88.6	35546	3	AC074202	AC074202 Leishmani
3	39	88.6	110000	1	BX571965_13	Continuation (14 0
4	39	88.6	110000	1	CP000010_19	Continuation (20 0

c	5	39	88.6	117146	2	AC087796	AC087796 Leishmani
c	6	39	88.6	142401	8	AC118669	AC118669 Genomic s
c	7	39	88.6	150106	2	AC144913	AC144913 Mus muscu
c	8	39	88.6	160650	3	AC008188	AC008188 Drosophil
c	9	39	88.6	160885	2	AC018184	AC018184 Drosophil
c	10	39	88.6	167588	8	AF461424	AF461424 Oryza sat
c	11	39	88.6	196178	10	AC107764	AC107764 Mus muscu
c	12	39	88.6	254192	2	AC146955	AC146955 Otolemur
c	13	39	88.6	221049	2	AC137178	AC137178 Rattus no
c	14	39	88.6	221586	2	BX957332	BX957332 Danio rer
c	15	39	88.6	238103	3	AC099140	AC099140 Rattus no
c	16	39	88.6	246934	2	AC094675	AC094675 Rattus no
c	17	39	88.6	306854	3	AE003797	AE003797 Drosophil
c	18	39	88.6	951	3	AY486427	AY486427 Trichinel
c	19	39	86.4	1963	10	M73991	M73991 Mesoricetu
c	20	38	86.4	38200	6	HAMALINH3	HAMALINH3
c	21	38	86.4	60940	6	AC000046	AC000046 Homo sapi
c	22	38	86.4	11990	3	AX695473	AX695473 Sequence
c	23	38	86.4	120472	9	AC091866	AC091866 Homo sapi
c	24	38	86.4	143298	8	AC091834	AC091834 Homo sapi
c	25	38	86.4	146438	8	AC134240	AC134240 Oryza sat
c	26	38	86.4	152788	3	AL117329	AL117329 Human DNA
c	27	38	86.4	157928	10	AC120512	AC120512 Rattus no
c	28	38	86.4	157944	5	AC112256	AC112256 Mus muscu
c	29	38	86.4	159717	2	AC092532	AC092532 Papio anu
c	30	38	86.4	160584	2	AC008368	AC008368 Trypanoso
c	31	38	86.4	162045	2	AC091655	AC091655 Trypanoso
c	32	38	86.4	170806	2	AC023488	AC023488 Trypanoso
c	33	38	86.4	177312	5	CR627483	CR627483 Danio rer
c	34	38	86.4	179599	2	BX005140	BX005140 zebrafish
c	35	38	86.4	185315	2	AC112930	AC112930 Mus muscu
c	36	38	86.4	187939	5	AC110188	AC110188 Mus muscu
c	37	38	86.4	190201	2	AL928672	AL928672 Zebrafish
c	38	38	86.4	195685	2	AC129837	AC129837 Papio anu
c	39	38	86.4	195690	2	AC119570	AC119570 Papio anu
c	40	38	86.4	196078	10	AC116045	AC116045 Papio ham
c	41	38	86.4	197796	2	AL663091	AL663091 Mouse DNA
c	42	38	86.4	202399	9	CR627483	CR627483 Danio rer
c	43	38	86.4	203399	9	AC129792	AC129792 Rattus no
c	44	38	86.4	205928	9	AC090966	AC090966 Papio anu
c	45	38	86.4	209060	9	AC010733	AC010733 Homo sapi

ALIGNMENTS

RESULT 1
AC113611

LOCUS   AC113611   60597 bp   DNA   linear   PRI 10-MAR-2003
DEFINITION Homo sapiens BAC clone RP11-421M20 from 4, complete sequence.
ACCESSION   AC113611
VERSION   AC113611.3 GI:28475761
KEYWORDS   HTG;
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 60597)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2 (bases 1 to 60597)
  AUTHORS   Isak,A. and Cotton,M.
  TITLE   The sequence of Homo sapiens BAC clone RP11-421M20
  JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 60597)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 60597)

```
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (11-APR-2002) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  5  (bases 1 to 60597)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (23-FEB-2003) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  6  (bases 1 to 60597)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (10-MAR-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT    On Feb 23, 2003 this sequence version replaced gi:20136967.
           -------------- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@watson.wustl.edu
           --------------   Summary Statistics
           Center project name: H_NH0421M20
           --------------.

NOTICE:    This sequence may not represent the entire insert of this
           clone. It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
This sequence is not the entire insert of the clone. This clone is
overlapped by AC104650 and AC104825.

Discrepant bases between AC104650, AC104825 and clone sequence.

Data from AC104825 was used to finish this clone.
```

FEATURES   Location/Qualifiers
     source   1..60597
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-421M20"
                     /clone_lib="RPCI-11"

     repeat_region   539..903
                     /rpt_family="ERVL"

     repeat_region   2797..2907
                     /rpt_family="(CAGAGA)n"
     repeat_region   2908..3114
                     /rpt_family="ERV1"
     repeat_region   3122..3184
                     /rpt_family="(CAGAGA)n"
     repeat_region   3256..3368
                     /rpt_family="MIR"
     repeat_region   3500..3682
                     /rpt_family="L1"
     repeat_region   3713..3782
                     /rpt_family="L1"
     repeat_region   5208..5410
                     /rpt_family="AT_rich"
     repeat_region   5451..5503
                     /rpt_family="MIR"
     repeat_region   5504..5677
                     /rpt_family="ERV1"
     repeat_region   6645..6687
                     /rpt_family="MIR"
     repeat_region   6998..7067
                     /rpt_family="MIR"
     repeat_region   7741..8173
                     /rpt_family="MIR"
     repeat_region   8530..8610
                     /rpt_family="MER1_type"
     misc_feature   8638..9514
                     /note="CpG island (%GC=76.7, o/e=0.80, #CpGs=114)"
     repeat_region   9027..9048
                     /rpt_family="GC_rich"
     repeat_region   9063..9111
                     /rpt_family="GC_rich"
     repeat_region   9612..9739
                     /rpt_family="MIR"
     repeat_region   10414..10509
                     /rpt_family="MIR"
     repeat_region   10688..11034
                     /rpt_family="MER103"
     repeat_region   11781..11997
                     /rpt_family="MaLR"
     repeat_region   12212..12325
                     /rpt_family="L2"
     repeat_region   13275..13625
                     /rpt_family="MIR"
     repeat_region   13703..13769
                     /rpt_family="L2"
     repeat_region   14926..15054
                     /rpt_family="CR1"
     repeat_region   15239..15281
                     /rpt_family="MIR"
     repeat_region   15349..15515
                     /rpt_family="Tip100"
     repeat_region   16000..16113
                     /rpt_family="MIR"
     repeat_region   16639..16617
                     /rpt_family="MIR"
     repeat_region   17127..17228
                     /rpt_family="L2"
     repeat_region   17883..17982
                     /rpt_family="Tip100"
     repeat_region   18108..18173
                     /rpt_family="MIR"
     repeat_region   19059..19127
                     /rpt_family="CT-rich"
     repeat_region   20087..20234
                     /rpt_family="L2"
     repeat_region   20576..20738
                     /rpt_family="MIR"
     repeat_region   22242..22380
                     /rpt_family="MIR"
     repeat_region   23733..23804
                     /rpt_family="(TGAA)n"
     repeat_region   23890..23987
                     /rpt_family="L2"

```
repeat_region      24011..24440     /rpt_family="MIR"
repeat_region      25254..25412     /rpt_family="MaLR"
repeat_region      26001..26150     /rpt_family="Alu"
repeat_region      26473..26512     /rpt_family="MIR"
repeat_region      26513..26894     /rpt_family="MIR"
repeat_region      26695..27048     /rpt_family="MaLR"
repeat_region      27423..27724     /rpt_family="MIR"
repeat_region      27725..28016     /rpt_family="Alu"
repeat_region      28029..28096     /rpt_family="Alu"
repeat_region      28099..28399     /rpt_family="(TTCC)n"
repeat_region      28406..28597     /rpt_family="Alu"
repeat_region      29613..29783     /rpt_family="MER1_type"
repeat_region      29988..30078     /rpt_family="L1"
repeat_region      30801..30893     /rpt_family="MIR"
```

Alignment Scores:
```
Pred. No.:
Score:                  491
Percent Similarity:     44.00          Matches:          60597
Best Local Similarity:  100.00%        Conservative:     0
Query Match:            100.00%        Mismatches:       0
DB:                     100.00%        Indels:           0
                        9              Gaps:             0
```

US-10-617-443B-2_COPY_1_9 (1-9) x AC113611 (1-60597)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
         ||||||||||||||||||||||||||||
Db   21559 ATGCACCTGGCCCTTCCCGCCAGCGCA 21585

RESULT 2
LOCUS   AC074202   35546 bp   DNA   linear   INV 28-JUN-2002
DEFINITION Leishmania major chromosome 35 clone L4946 strain Friedlin,
            complete sequence.
ACCESSION   AC074202
VERSION   AC074202.5   GI:21622732
KEYWORDS   HTG.
SOURCE   Leishmania major
  ORGANISM Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1 (bases 1 to 35546)
  AUTHORS   Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
            McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-JUL-2000) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE   2 (bases 1 to 35546)
  AUTHORS   Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
            McDonagh,P., Ivens,A., Nguyen,D., Munden,H., Stuart,K., Nelson,S.,
            Marty,A., Seyler,A. and Mack,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE   3 (bases 1 to 35546)
  AUTHORS   Worthey,E.A., Sisk,E., Seyler,A., Nelson,S., Vogt,C., Robertson,L.,
            Pentony,M., Ivens,A., Mack,J., Munden,H., Marty,A., Stuart,K. and
            Myler,P.J.
  TITLE   Direct Submission
```


```
  JOURNAL   Submitted (21-JUN-2001) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE   4  (bases 1 to 35546)
  AUTHORS   Worthey,E.A., Sisk,E., Seyler,A., Nelson,S., Vogt,C., Robertson,L.,
            Pentony,M., Ivens,A., Mack,J., Munden,H., Marty,A., Stuart,K. and
            Myler,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-2001) Seattle Biomedical Research Institution,
            Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE   5  (bases 1 to 35546)
  AUTHORS   Worthey,E.A., Sisk,E., Seyler,A., Nelson,S., Vogt,C., Robertson,L.,
            Pentony,M., Ivens,A., Mack,J., Munden,H., Marty,A., Stuart,K. and
            Myler,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2002) Seattle Biomedical Research Institution, USA
COMMENT     On Jun 28, 2002 this sequence version replaced gi:16751888.
FEATURES             Location/Qualifiers
     source          1..35546
                     /organism="Leishmania major"
                     /mol_type="genomic DNA"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /chromosome="35"
                     /clone="L4946"
ORIGIN
```

Alignment Scores:
```
Pred. No.:
Score:                  3.73e+03
Percent Similarity:     39.00          Matches:          8
Best Local Similarity:  88.89%         Conservative:     0
Query Match:            88.89%         Mismatches:       1
DB:                     88.64%         Indels:           1
                        3              Gaps:             0
                                       Length:           35546
```

US-10-617-443B-2_COPY_1_9 (1-9) x AC074202 (1-35546)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
         ||||||||||||||||||||||||||||
Db 966 ATGCACCTGCGCCACACCTGCATCCGCT 992

RESULT 3
LOCUS   BX571965 13
WPCOMMENT
COMMENT   Sequence split into 41 fragments LOCUS BX571965 Accession BX571965

Fragment Name	Begin	End
BX571965_00	1	110000
BX571965_01	100001	210000
BX571965_02	200001	310000
BX571965_03	300001	410000
BX571965_04	400001	510000
BX571965_05	500001	610000
BX571965_06	600001	710000
BX571965_07	700001	810000
BX571965_08	800001	910000
BX571965_09	900001	1010000
BX571965_10	1000001	1110000
BX571965_11	1100001	1210000
BX571965_12	1200001	1310000
BX571965_13	1300001	1410000
BX571965_14	1400001	1510000
BX571965_15	1500001	1610000
BX571965_16	1600001	1710000
BX571965_17	1700001	1810000
BX571965_18	1800001	1910000
BX571965_19	1900001	2010000
BX571965_20	2000001	2110000
BX571965_21	2100001	2210000
BX571965_22	2200001	2310000
BX571965_23	2300001	2410000
BX571965_24	2400001	2510000
BX571965_25	2500001	2610000
BX571965_26	2600001	2710000
```

1029

Continuation (14 of 41) of BX571965 from base 1300001 (BX571965 Burkholderia pseudomalle...

| | Begin | End |
|---|---|---|
| BX571965_27 | 2700001 | 2810000 |
| BX571965_28 | 2800001 | 2910000 |
| BX571965_29 | 2900001 | 3010000 |
| BX571965_30 | 3000001 | 3110000 |
| BX571965_31 | 3100001 | 3210000 |
| BX571965_32 | 3200001 | 3310000 |
| BX571965_33 | 3300001 | 3410000 |
| BX571965_34 | 3400001 | 3510000 |
| BX571965_35 | 3500001 | 3610000 |
| BX571965_36 | 3600001 | 3710000 |
| BX571965_37 | 3700001 | 3810000 |
| BX571965_38 | 3800001 | 3910000 |
| BX571965_39 | 3900001 | 4010000 |
| BX571965_40 | 4000001 | 4074542 |

Alignment Scores:
Pred. No.: 1
Score: 1.04e+04
Percent Similarity: 39.00
Best Local Similarity: 88.89%
Query Match: 88.89%
DB: 1

Length: 110000
Matches: 8
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BX571965_13 (1-110000)

Qy      1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||
Db  53646 ATGCATCTGGCCACGCCCGGTCCGCG 53672

RESULT 4
WPCOMMENT
CP000010_19/c
Sequence split into 36 fragments    LOCUS CP000010 Accession CP000010

| Fragment Name | Begin | End |
|---|---|---|
| CP000010_00 | 1 | 110000 |
| CP000010_01 | 100001 | 210000 |
| CP000010_02 | 200001 | 310000 |
| CP000010_03 | 300001 | 410000 |
| CP000010_04 | 400001 | 510000 |
| CP000010_05 | 500001 | 610000 |
| CP000010_06 | 600001 | 710000 |
| CP000010_07 | 700001 | 810000 |
| CP000010_08 | 800001 | 910000 |
| CP000010_09 | 900001 | 1010000 |
| CP000010_10 | 1000001 | 1110000 |
| CP000010_11 | 1100001 | 1210000 |
| CP000010_12 | 1200001 | 1310000 |
| CP000010_13 | 1300001 | 1410000 |
| CP000010_14 | 1400001 | 1510000 |
| CP000010_15 | 1500001 | 1610000 |
| CP000010_16 | 1600001 | 1710000 |
| CP000010_17 | 1700001 | 1810000 |
| CP000010_18 | 1800001 | 1910000 |
| CP000010_19 | 1900001 | 2010000 |
| CP000010_20 | 2000001 | 2110000 |
| CP000010_21 | 2100001 | 2210000 |
| CP000010_22 | 2200001 | 2310000 |
| CP000010_23 | 2300001 | 2410000 |
| CP000010_24 | 2400001 | 2510000 |
| CP000010_25 | 2500001 | 2610000 |
| CP000010_26 | 2600001 | 2710000 |
| CP000010_27 | 2700001 | 2810000 |
| CP000010_28 | 2800001 | 2910000 |
| CP000010_29 | 2900001 | 3010000 |
| CP000010_30 | 3000001 | 3110000 |
| CP000010_31 | 3100001 | 3210000 |
| CP000010_32 | 3200001 | 3310000 |
| CP000010_33 | 3300001 | 3410000 |
| CP000010_34 | 3400001 | 3510000 |
| CP000010_35 | 3500001 | 3510148 |

Continuation (20 of 36) of CP000010 from base 1900001 (CP000010 Burkholderia mallei ATCC...

Alignment Scores:
Pred. No.: 1
Score: 1.04e+04
Percent Similarity: 39.00
Best Local Similarity: 88.89%
Query Match: 88.89%
DB: 1

Length: 110000
Matches: 8
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x CP000010_19 (1-110000)

Qy      1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||
Db  65822 ATGCATCTGGCCACGCCCGGTCCGCG 65796

RESULT 5
AC087796/c
LOCUS       AC087796               117146 bp    DNA     linear   HTG 14-APR-2003
DEFINITION  Leishmania major chromosome 35 clone P568 strain Friedlin, ***
            SEQUENCING IN PROGRESS ***.
ACCESSION   AC087796
VERSION     AC087796.4  GI:18158370
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Leishmania major
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 117146)
  AUTHORS   Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
            McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2001) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE   2  (bases 1 to 117146)
  AUTHORS   Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
            McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2003) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT     On Jan 16, 2002 this sequence version replaced gi:15778699.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            * 117146: contig of 117146 bp in length.
FEATURES             Location/Qualifiers
     source          1..117146
                     /organism="Leishmania major"
                     /mol_type="genomic DNA"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /chromosome="35"
                     /clone="P568"
ORIGIN

Alignment Scores:
Pred. No.: 1
Score: 1.1e+04
Percent Similarity: 39.00
Best Local Similarity: 88.89%
Query Match: 88.89%
DB: 2

Length: 117146
Matches: 8
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x AC087796 (1-117146)

Qy      1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||
Db  29532 ATGCACCTGGCCACCTGCATCCGCT 29506

## RESULT 6

```
LOCUS       AC118669       142401 bp    DNA     linear   PLN 04-SEP-2002
DEFINITION  Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBb0011G21, from chromosome 3, complete sequence.
ACCESSION   AC118669
VERSION     AC118669.2  GI:22711575
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 142401)
AUTHORS     McCombie,W.R., de la Bastide,M., Spiegel,L., Preston,R.,
            Nascimento,L., Zutavern,T., Balija,V., Bell,M., Miller,B.,
            Katzenberger,F., Muller,S., Sullivan,P., Yang,C., Dike,S.,
            O'Shaughnessy,A., Palmer,L. and Dedhia,N.
TITLE       Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBb0011G21, from chromosome 3, complete sequence
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 142401)
AUTHORS     McCombie,W.R.
TITLE       Direct Submission
JOURNAL     Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   3  (bases 1 to 142401)
AUTHORS     McCombie,W.R.
TITLE       Direct Submission
JOURNAL     Submitted (20-APR-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Sep 4, 2002 this sequence version replaced gi:20219081.
            Submitted (04-SEP-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The nucleotide
            sequence of this BAC clone was generated by combining Syngenta and
            Cold Spring Harbor Laboratory Genome Center sequencing data.
            Clone OSJNBb0011G21 overlaps clone OSJNBb0008SH08 (AC118340) from
            base 67378 to base 142401. The overlap is from base 52004 to base
            126708 (complemented) on OSJNBb0008SH08.
FEATURES             Location/Qualifiers
     source          1..142401
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="3"
                     /clone="OSJNBb0011G21"
                     /clone_lib="EcoRI"
     misc_feature    1585..11608
                     /note="We believe the assembly to be correct. The
                     sequence is a mononucleotide (C) repeat in which the exact
                     number of Cs is unknown. One subclone spans the repeat
                     into unique sequence on both sides."
     misc_feature    66157..66177
                     /note="We believe the assembly to be correct. The
                     sequence is a mononucleotide (G) repeat in which the exact
                     number of Gs is unknown. Four subclones in the region
                     show one additional G compared to that represented by the
                     assembly."
ORIGIN
```

Alignment Scores:

Pred. No.:	1.31e+04
Score:	39.00
Percent Similarity:	100.00%
Best Local Similarity:	77.78%

Length:	142401
Matches:	7
Conservative:	2
Mismatches:	0

RESULT 7

```
Query Match:             88.64%        Indels:        0
DB:         8                          Gaps:          0

US-10-617-443B-2_COPY_1_9 (1-9) x AC118669 (1-142401)

Qy         1 MetHisLeuAlaLeuProAlaSerAla 9
             :::|||||||:::|||||::||||||||||
Db     33763 CTCCACCTTAGCCATGCCGCCGCCTCCGCG 33737
```

```
LOCUS       AC144913       150106 bp    DNA     linear   HTG 21-JUL-2003
DEFINITION  Mus musculus chromosome 15 clone RP24-203A19 map 15, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC144913
VERSION     AC144913.2  GI:33086398
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 150106)
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE       Direct Submission
JOURNAL     Mus musculus chromosome 15, clone RP24-203A19
REFERENCE   2  (bases 1 to 150106)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,M., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (25-MAY-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 150106)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Lander,E., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,M., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (21-JUL-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

COMMENT   On Jul 21, 2003 this sequence version replaced gi:31074896.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------------- Project Information
            Center project name: L29103
            Center clone name: 203_A_19

            Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 142901 bases at least Q40
            Consensus quality: 146204 bases at least Q30
            Consensus quality: 147399 bases at least Q20
            Insert size: 127000; agarose-fp
            Insert size: 148506; sum-of-contigs
            Quality coverage: 10.9 in Q20 bases; agarose-fp
            Quality coverage: 9.3 in Q20 bases; sum-of-contigs

      * NOTE: This is a 'working draft' sequence. It currently
      * consists of 17 contigs. The true order of the pieces
      * is not known and their order in this sequence record is
      * arbitrary. Gaps between the contigs are represented as
      * runs of N, but the exact sizes of the gaps are unknown.
      * This record will be updated with the finished sequence
      * as soon as it is available and the accession number will
      * be preserved.
      *
      *   1   15613: contig of 15613 bp in length
      *   15614   15713: gap of 100 bp
      *   15714   16443: contig of 730 bp in length
      *   16444   16543: gap of 100 bp
      *   16544   17157: contig of 614 bp in length
      *   17158   17257: gap of 100 bp
      *   17258   18600: contig of 1343 bp in length
      *   18601   18700: gap of 100 bp
      *   18701   19955: contig of 1255 bp in length
      *   19956   20055: gap of 100 bp
      *   20056   21611: contig of 1556 bp in length
      *   21612   21711: gap of 100 bp
      *   21712   23269: contig of 1558 bp in length
      *   23270   23369: gap of 100 bp
      *   23370   25160: contig of 1791 bp in length
      *   25161   25260: gap of 100 bp
      *   25261   27996: contig of 2736 bp in length
      *   27997   28096: gap of 100 bp
      *   28097   30632: contig of 2536 bp in length
      *   30633   30732: gap of 100 bp
      *   30733   37468: contig of 6736 bp in length
      *   37469   37568: gap of 100 bp
      *   37569   83287: contig of 45719 bp in length
      *   83288   83387: gap of 100 bp
      *   83388   93357: contig of 9970 bp in length
      *   93358   93457: gap of 100 bp
      *   93458 108049: contig of 14592 bp in length
      * 108050 108149: gap of 100 bp
      * 108150 122020: contig of 13871 bp in length
      * 122021 122120: gap of 100 bp
      * 122121 141979: contig of 19859 bp in length
      * 141980 142079: gap of 100 bp
      * 142080 150106: contig of 8027 bp in length.

FEATURES   Location/Qualifiers
     source   1..150106
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP24-203A19"

     misc_feature   /clone_lib="RPCI-24 Male Mouse BAC"
                     1..15613
                     /note="assembly_fragment
                     vector_end:SP6
                     clone_end:SP6"
     misc_feature   15714..16443
                     /note="assembly_fragment
                     clone_end:left"
     misc_feature   16544..17157
                     /note="assembly_fragment"
     misc_feature   17258..18600
                     /note="assembly_fragment"
     misc_feature   18701..19955
                     /note="assembly_fragment"
     misc_feature   20056..21611
                     /note="assembly_fragment"
     misc_feature   21712..23269
                     /note="assembly_fragment"
     misc_feature   23370..25160
                     /note="assembly_fragment"
     misc_feature   25261..27996
                     /note="assembly_fragment"
     misc_feature   28097..30632
                     /note="assembly_fragment"
     misc_feature   30733..37468
                     /note="assembly_fragment"
     misc_feature   37569..83287
                     /note="assembly_fragment"
     misc_feature   83388..93357
                     /note="assembly_fragment"
     misc_feature   93458..108049
                     /note="assembly_fragment"
     misc_feature   108150..122020
                     /note="assembly_fragment"
     misc_feature   122121..141979
                     /note="assembly_fragment"
     misc_feature   142080..150106
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

US-10-617-443B-2_COPY_1_9 (1-9) x AC144913 (1-150106)

Qy   2 HisLeuAlaLeuProAlaSerAla 9
             ||||||||||||||||||||||||
Db   44486 CATCTTGCTTCCTGCCTCGGCC 44509

Alignment Scores:
Pred. No.:   1
Score: 1.38e+04
Percent Similarity: 39.00
Best Local Similarity: 100.00%
Query Match: 88.64%
DB: 2

Length: 150106
Matches: 8
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

RESULT 8
AC008188/c
LOCUS   AC008188   160650 bp   DNA   linear   INV 27-APR-2001
DEFINITION   Drosophila melanogaster, chromosome 2R, region 56A2-56B1, BAC clone
             BACR08I18, complete sequence.
ACCESSION   AC008188
VERSION   AC008188.4 GI:1381832
KEYWORDS   HTG.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 160650)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
             Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
             Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
             Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,

(continuation of preceding RESULT — AC008188, Drosophila)

REFERENCE
AUTHORS   Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
   Ferriera,S., Priae,E., Galle,R.F., Garg,N.S., George,R.A.,
   Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
   Ibegwan,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
   McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
   Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
   Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
   Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
   Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE   Sequencing of Drosophila chromosome 2R, region 56A2-56B1
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 160650)
AUTHORS   Celniker,S.E., Agbayani,A., Araina,T.T., Baxter,E., Blazej,R.G.,
   Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,N.,
   Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
   Hoskins,R.A., Houston,K.A., Hunmasti,S.R., Karra,K., Kearney,L.,
   Kim,E., Lee,B., Lewis,S., Li,P., Lomctan,M.A., Mazda,P.,
   Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
   Pfeiffer,B., Poon,D., Sequeira,A., Sethi,H., Snir,E.,
   Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
   Rubin,G.M.
TITLE   Direct Submission
JOURNAL   Submitted (29-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
   Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT   On Apr 27, 2001 this sequence version replaced gi:6598785.
   Sequence submitted by:
   Berkeley Drosophila Genome Project
   Lawrence Berkeley National Laboratory, MS 64-121
   Berkeley, CA 94720
   This sequence was assembled using end sequences from a whole genome
   shotgun and from subclones of this BAC and its neighboring clones.
   For further information about this sequence, including its location
   and relationship to other sequences, please visit our sequence
   archive web site (http://www.fruitfly.org/sequence/) or send email
   to bdgp@fruitfly.berkeley.edu.
FEATURES   Location/Qualifiers
   source   1..160650
     /organism="Drosophila melanogaster"
     /mol_type="genomic DNA"
     /strain="y; cn bw sp"
     /chromosome="2R"
     /map="56A2-56B1"
     /clone="BACR08I18 (D660)"
     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
     Drosophila melanogaster BAC library, partial EcoRI in
     pBACe3.6)"
ORIGIN

US-10-617-443B-2_COPY_1_9 (1-9) x AC008188 (1-160650)

Alignment Scores:

Pred. No.:	1.46e+04
Score:	39.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	88.64%
DB:	3
Length:	160650
Matches:	8
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

```
Db   81657  CACTTGGCATTGCCAGCCAGCGCC  81634
                    ||||||||||||||||||||||||
Qy       2  HisLeuAlaLeuProAlaSerAla  9
```

RESULT 9

LOCUS   AC018184   160885 bp   DNA   linear   HTG   09-DEC-1999
DEFINITION   Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION   AC018184
VERSION   AC018184.1   GI:6553007
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
   Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 160885)
AUTHORS   Adams,M. and Venter,J.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
   Rockville, MD, USA
COMMENT   This sequence was identified as CDM:10213716 by the submitter.
   For more information on this record e-mail to fly@celera.com.
   * NOTE: This sequence is a 'working draft' sequence.
   * This sequence will be replaced
   * by the finished sequence as soon as it is available and
   * the accession number will be preserved.
FEATURES   Location/Qualifiers
   source   1..160885
     /organism="Drosophila melanogaster"
     /mol_type="genomic DNA"
     /db_xref="taxon:7227"
ORIGIN

US-10-617-443B-2_COPY_1_9 (1-9) x AC018184 (1-160885)

Alignment Scores:

Pred. No.:	1.47e+04
Score:	39.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	88.64%
DB:	2
Length:	160885
Matches:	8
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

```
Db  125771  CACTTGGCATTGCCAGCCAGCGCC  125748
                    ||||||||||||||||||||||||
Qy       2  HisLeuAlaLeuProAlaSerAla  9
```

(handwritten annotation: 2/27   -86.7%)

RESULT 10

LOCUS   AF461424   167588 bp   DNA   linear   PLN   26-JUN-2003
DEFINITION   Oryza sativa chromosome 3 clone OSJNBa0016B07, complete sequence.
ACCESSION   AF461424
VERSION   AF461424.2   GI:32265039
KEYWORDS   HTG.
SOURCE   Oryza sativa
ORGANISM   Oryza sativa
   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
   Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 167588)
AUTHORS   Eastman,A.P., Smith,S.C., Gingle,A.R., Pratt,L.H. and
   Cordonnier-Pratt,M.-M.
TITLE   Direct Submission
JOURNAL   Submitted (18-DEC-2001) Plant Sciences, Athens, University of Georgia, 3604 Miller
   Plant Sciences, Athens, GA 30606, USA
REFERENCE   2 (bases 1 to 167588)
AUTHORS   Eastman,A.P., Smith,S.C., Bertin,N., Liang,C., Gingle,A.R.,
   Pratt,L.H. and Cordonnier-Pratt,M.-M.
TITLE   Direct Submission
JOURNAL   Submitted (26-JUN-2003) Botany, University of Georgia, 3604 Miller
   Plant Sciences, Athens, Georgia 30606, USA
REMARK   Sequence update by submitter
COMMENT   On Jun 26, 2003 this sequence version replaced gi:18150816.
FEATURES   Location/Qualifiers
   source   1..167588
     /organism="Oryza sativa"
     /mol_type="genomic DNA"
     /db_xref="taxon:4530"
     /chromosome="3"
     /clone="OSJNBa0016B07"
ORIGIN

US-10-617-443B-2_COPY_1_9 (1-9) x AF461424 (1-167588)

Alignment Scores:

Pred. No.:	1.52e+04
Score:	39.00
Percent Similarity:	100.00%
Length:	167588
Matches:	7
Conservative:	2

RESULT 11

US-10-617-443B-2_COPY_1_9 (1-9) x AF461424 (1-167588)

```
Qy          1 MetHisLeuAlaLeuProAlaSerAla 9
              ::|||||||||:|:|||||:||||||||
Db      12713 CTCCACCTAGCCATGCCTCGCCCCTCCGCG 12739
```

```
US-10-617-443B-2_copy_1_9 (1-167588)

Db    12713 CTCCACCTAGCCATGCCTCGCCCCTCCGCG 12739
```

LOCUS   AC107764   196178 bp   DNA   linear   ROD 07-JUL-2004
DEFINITION Mus musculus chromosome 15, clone RP23-179A6, complete sequence.
ACCESSION   AC107764
VERSION   AC107764.11 GI:49745564
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 196178)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Mus musculus chromosome 15, clone RP23-179A6
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 196178)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Illev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,R., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,I., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,J.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,K., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 196178)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Illev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,J., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-MAY-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 196178)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Illev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-JUL-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jul 7, 2004 this sequence version replaced gi:47106120.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
            Center: Whitehead Institute/MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
            Center project name: L20413
            Center clone name: 179_A_6
FEATURES   Location/Qualifiers
     source   1..196178
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP23-179A6"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     repeat_region   1..84
                     /rpt_family="B1_MM"
     misc_feature   1..6
                     /note="clone_boundary
                     clone_end:SP6
                     site:EcoRI"
     repeat_region   102..213
                     /rpt_family="PB1D7"
     repeat_region   214..255
                     /rpt_family="PB1D7"
     repeat_region   265..428
                     /rpt_family="B3"
     repeat_region   429..475
                     /rpt_family="B3"
     repeat_region   1982..2027
                     /rpt_family="PB1D7"
     repeat_region   complement(3183..3312)
                     /rpt_family="B1F"
     unsure   7340..7423
                     /note="single clone coverage"
     unsure   8810..8925
                     /note="single clone coverage"
     repeat_region   9195..9262
                     /rpt_family="B1F"

```
repeat_region    /note="single clone coverage"
                 10007..10115
repeat_region    complement(11227..11336)
                 /rpt_family="B1_MM"
repeat_region    complement(11361..11471)
                 /rpt_family="PB1D9"
repeat_region    11798..11847
                 /rpt_family="MIR"
repeat_region    /rpt_family="MIR3"
                 13396..13475
repeat_region    complement(12475..12582)
                 /rpt_family="B1_MM"
repeat_region    complement(12632..12760)
                 /rpt_family="PB1D7"
repeat_region    complement(13039..13137)
                 /rpt_family="PB1D7"
repeat_region    13671..13744
                 /rpt_family="(TA)n"
repeat_region    13745..13764
                 /rpt_family="(TG)n"
repeat_region    complement(13815..13903)
                 /rpt_family="(CATATA)n"
repeat_region    complement(13830..13922)
                 /rpt_family="ID_B1"
repeat_region    complement(14135..14866)
                 /rpt_family="LIME2"
repeat_region    complement(14902..15095)
                 /rpt_family="B2_Mm2"
repeat_region    15162..15213
                 /rpt_family="(TATG)n"
repeat_region    complement(15312..15477)
                 /rpt_family="RSINE1"
repeat_region    complement(15478..15748)
                 /rpt_family="LIME2"
repeat_region    complement(15779..15864)
                 /rpt_family="L3"
repeat_region    complement(15895..16043)
                 /rpt_family="B1_MM"
repeat_region    16082..16620
                 /rpt_family="B1_MM"
repeat_region    complement(16327..16466)
                 /rpt_family="AT_rich"
repeat_region    complement(16615..16735)
                 /rpt_family="PB1D10"
repeat_region    complement(16824..16965)
                 /rpt_family="ID_B1"
repeat_region    16968..16991
                 /rpt_family="B3"
repeat_region    complement(17450..17553)
                 /rpt_family="PB1D10"
repeat_region    complement(18838..18358)
                 /rpt_family="B4"
repeat_region    21799..21898
                 /rpt_family="B1_MM"
repeat_region    24013..24050
                 /rpt_family="B4A"
repeat_region    complement(23074..23267)
                 /rpt_family="PB1D10"
repeat_region    complement(23435..23562)
                 /rpt_family="B3A"
repeat_region    complement(23437..23598)
                 /rpt_family="B1_MM"
repeat_region    complement(23708..23884)
                 /rpt_family="ID_B1"
repeat_region    24287..24322
                 /rpt_family="(CTG)n"
```

```
DB:

RESULT 12
AC146955/c

LOCUS       AC146955    218827 bp    DNA    linear    HTG 30-OCT-2003
DEFINITION  Otolemur garnettii clone CH256-263M9, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC146955
VERSION     AC146955.1  GI:38044153
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Otolemur garnettii (small-eared galago)
  ORGANISM  Otolemur garnettii
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Otolemur.
REFERENCE   1  (bases 1 to 218827)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
            Coleman,H., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Han,J., Hansen,N., Ho,S-L, Hu,P., Hurle,B., Idol,J.R.,
            Karlins,E., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Marguiles,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K.,
            Schueler,M.G., Shah,K., Sison,C., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and
            Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218827)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     ------------  Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            ------------  Project Information
            Center project name: fkv
            Center clone name: 263M09
            ------------  Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 213879 bases at least Q40
            Consensus quality: 215076 bases at least Q30
            Consensus quality: 215592 bases at least Q20
            Insert size: 234000; agarose-fp
            Insert size: 217327; sum-of-contigs
            Quality coverage: 9.25x in Q20 bases; agarose-fp
            Quality coverage: 9.96x in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *     1   2402: contig of 2402 bp in length
            *  2403   2502: gap of unknown length
            *  2503   9514: contig of 7012 bp in length
            *  9514   9614: gap of unknown length
            *  9615  14216: contig of 4602 bp in length
            * 14217  14316: gap of unknown length
```

```
US-10-617-443B-2_COPY_1_9 (1-9) x AC107764 (1-196178)

                        10                                    Gaps:  0

Qy    2 HisLeuAlaLeuProAlaSerAla 9
        |||||||||||||||||||||||||
Db  263 CATCTTGCTTCTTCCTCGCCTCGGCC 240
```

```
Alignment Scores:
Pred. No.:              1.76e+04        Length:        196178
Score:                  39.00           Matches:       8
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            88.64%          Indels:        0
```

```
US-10-617-443B-2_COPY_1_9 (1-9) x AC146955 (1-218827)

Qy        2 HisLeuAlaLeuProAlaSerAla 9
            ||||||||||||||||||||||||
Db   182222 CATCTGGGCTCTTCCAGCCCTCTGCC 182199

RESULT 13
AC137178/c
LOCUS           AC137178             221049 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION      Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
                ***, 10 unordered pieces.
ACCESSION       AC137178
VERSION         AC137178.1  GI:25073057
KEYWORDS        HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE          Rattus norvegicus (Norway rat)
  ORGANISM      Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
REFERENCE    1  (bases 1 to 221049)
  AUTHORS       Muzny,D.,Marie,A., Metzker,M.,Lee,., Abramzon,S., Adams,C., Alder,J.,
                Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
                Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
                Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
                Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
                Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
                Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
                Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
                Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
                Davila,M.L., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
                Delgado,O., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
                Draper,H., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
                Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
                Fraser,C.M., Gabisi,A., Ganta,S., Garcia,A., Garner,T., Garza,M.,
                Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,C., Guevara,W.,
                Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
                Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
                Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
                Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,M.,
                Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
                Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
                Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
                Liu,J., Liu,W., Liu,J., London,P., Longacre,S., Lopez,J.,
                Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
                Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
                Mangun,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
                Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
                Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
                Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
                Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
                Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pai,S.,
                Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
                Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
                Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
                Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
                Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
                Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
                Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
                Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
                Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
                Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
                Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
                Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
                Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
                Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
                Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
                Niederhausern,A.C., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
                Weinstock,G. and Gibbs,R.A.
  TITLE         Direct Submission
  JOURNAL       Unpublished
REFERENCE    2  (bases 1 to 221049)
```

ORIGIN

```
Alignment Scores:
Pred. No.:                 1.94e+04
Score:                     39.00
Percent Similarity:        100.00%          Length:        218827
Best Local Similarity:     100.00%          Matches:            8
Query Match:               88.64%           Conservative:       0
DB:                        2                Mismatches:         0
                                            Indels:             0
                                            Gaps:               0
```

FEATURES   Location/Qualifiers
     source   1..218827
                     /organism="Otolemur garnettii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:30611"
                     /clone="CH256-263M9"
                     /clone_lib="CH256"
     misc_feature   1..2402
                     /note="assembly_fragment"
     misc_feature   2503..9514
     misc_feature   9615..14216
     misc_feature   14317..24704
                     /note="assembly_fragment"
     misc_feature   24805..31640
                     /note="assembly_fragment"
     misc_feature   31741..43787
                     /note="assembly_fragment"
     misc_feature   43888..51984
                     /note="assembly_fragment"
     misc_feature   52085..63590
                     /note="assembly_fragment"
     misc_feature   63691..73810
                     /note="assembly_fragment"
     misc_feature   73911..88004
                     /note="assembly_fragment"
     misc_feature   88105..107084
                     /note="assembly_fragment"
     misc_feature   107185..124320
                     /note="assembly_fragment"
     misc_feature   124421..145411
                     /note="assembly_fragment"
     misc_feature   145512..164902
                     /note="assembly_fragment"
     misc_feature   165003..189267
                     /note="assembly_fragment"
     misc_feature   189368..218827
                     /note="assembly_fragment"

 *   14317   24704: contig of 10388 bp in length
 *   24705   24804: gap of unknown length
 *   31640   31641: contig of 6836 bp in length
 *   31741   43787: gap of unknown length
 *   43787   43887: contig of 12047 bp in length
 *   43888   51984: gap of unknown length
 *   51984   51985: contig of 8097 bp in length
 *   52084   52085: gap of unknown length
 *   63590   63591: contig of 11506 bp in length
 *   63691   63690: gap of unknown length
 *   73810   73811: contig of 10120 bp in length
 *   73911   73910: gap of unknown length
 *   88004   88005: contig of 14094 bp in length
 *   88105   88104: gap of unknown length
 *   107084   107085: contig of 18980 bp in length
 *   107185   107184: gap of unknown length
 *   124320   124321: contig of 17136 bp in length
 *   124421   124420: gap of unknown length
 *   145411   145412: contig of 20991 bp in length
 *   145512   145511: gap of unknown length
 *   164902   164521: contig of 19391 bp in length
 *   165003   164902: gap of unknown length
 *   165002   165003: contig of 24265 bp in length
 *   189267   189267: gap of unknown length
 *   189368   189267: contig of 24465 bp in length
 *   218827   189367: gap of unknown length
 *   189368   218827: contig of 29460 bp in length.
```

AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT   The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each contig-scaffold,
           individual sequence contigs are ordered and oriented and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the ends of the clone and there may be sequence
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contig will be indicated in the feature
           table.
           ----------------------- Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu/
           Contact: hgsc-help@bcm.tmc.edu
           ----------------------- Project Information
           Center project name: K2SK
           Center clone name: CH230-unknown
           ----------------------- Summary Statistics
           Assembly program: Phrap; version 0.990329
           Consensus quality: 155203 bases at least Q40
           Consensus quality: 161146 bases at least Q30
           Consensus quality: 165262 bases at least Q20
           Estimated insert size: 159376; sum-of-contigs estimation
           Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
           ----
         * NOTE: Estimated insert size may differ from sequence length
         *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 10 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *
         *   1   10473: contig of 10473 bp in length
         *   10474   10573: gap of unknown length
         *   10574   182909: contig of 172336 bp in length
         * 182910   183009: gap of unknown length
         * 183010   188163: contig of 5154 bp in length
         * 188164   188263: gap of unknown length
         * 188264   204169: contig of 15906 bp in length
         * 204170   204269: gap of unknown length
         * 204270   209800: contig of 5531 bp in length
         * 209801   209900: gap of unknown length
         * 209901   215096: contig of 5196 bp in length
         * 215097   215196: gap of unknown length
         * 215197   216256: contig of 1060 bp in length
         * 216257   216356: gap of unknown length
         * 216357   217914: contig of 1558 bp in length
         * 217915   218014: gap of unknown length
         * 218015   219553: contig of 1539 bp in length
         * 219554   219653: gap of unknown length
         * 219654   221049: contig of 1396 bp in length

FEATURES   Location/Qualifiers
     source   1..221049
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-unknown"
     misc_feature   1483..16398
                     /note="wgs_contig"
     misc_feature   1649..17449
                     /note="wgs_contig"
     misc_feature   61410..63256
                     /note="wgs_contig"
     misc_feature   74577..75784
                     /note="wgs_contig"
     misc_feature   96819..99058
                     /note="wgs_contig"
     misc_feature   99103..101394
                     /note="wgs_contig"
     misc_feature   112096..113281
                     /note="wgs_contig"
     misc_feature   116970..119727
                     /note="wgs_contig"
     misc_feature   134730..136296
                     /note="wgs_contig"
     misc_feature   162057..163329
                     /note="wgs_contig"
     misc_feature   175836..177673
                     /note="wgs_contig"
     misc_feature   188264..189718
                     /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:   1.96e+04   Length: 221049
Score: 39.00   Matches: 8
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 88.64%   Indels: 0
DB: 2   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x AC137178 (1-221049)

Qy   2 HisLeuAlaLeuProAlaSerAla 9
             ||||||||||||||||||||||||
Db   54018 CATCTCGCCCTGCCAGCCTCAGCC 53995

RESULT 14
BX957332
BX957332
LOCUS   BX957332   221586 bp   DNA   linear   HTG 10-OCT-2004
DEFINITION Danio rerio clone DKEY-170K13, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   BX957332
VERSION   BX957332.4 GI:54021781
KEYWORDS   HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Danio rerio (zebrafish)
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 221586)
AUTHORS   Barker,G.
TITLE   Direct Submission
JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Oct 10, 2004 this sequence version replaced gi:52694723.
            ----------------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ----------------------- Project Information
            Center project name: zK170K13
            ----------------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 220388 bases at least Q40
            Consensus quality: 220594 bases at least Q30
            Consensus quality: 220727 bases at least Q20
            Insert size: 220086; sum-of-contigs
            Insert size: 220053; 2.6% error; agarose-fp
            Quality coverage: 7.93x in Q20 bases; sum-of-contigs Quality
            coverage: 8.01x in Q20 bases; agarose-fp

```
RESULT 15
LOCUS        AC099140            238103 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION   Rattus norvegicus clone CH230-52A14, WORKING DRAFT SEQUENCE, 4
             unordered pieces.
ACCESSION    AC099140
VERSION      AC099140.12  GI:30522132
KEYWORDS     HTG; HTGS_DRAFT; HTGS_PHASE1; HTGS_FULLTOP.
```

```
SOURCE       Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 238103)
  AUTHORS    Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
             Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
             Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
             Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
             Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
             Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
             Cardenas,V., Carter,K., Cavazos,I., Ceaaar,H., Center,E.,
             Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
             Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
             Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
             Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
             Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
             Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
             Fernandez,S., Finley,R., Flagg,N., Forbes,L., Foster,M., Foster,P.,
             Fraser,C.M., Gabiei,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
             Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
             Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,P.,
             Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
             Hernandez,R., Hines,S., Hladun,S., Hodgson,A., Hogues,M.,
             Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
             Jackson,L., Jacobi,J., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
             Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
             Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
             Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
             Lorensuhewa,L., Louiseged,H., Lozado,R., Lu,X., Ma,J.,
             Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,
             Mangun,B., Mapua,P., Martin,M., Martin,R., Martinez,E.,
             Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,M.,
             Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
             Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
             Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
             Nwaokelemeh,O., Okwuonu,G., Olairpunsagoon,A., Pal,S.,
             Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
             Plopper,F., Poindexter,H., Popovic,D., Primus,E., Pu,L-L.,
             Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
             Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
             Rives,J., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
             Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
             Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smejs,D.,
             Sneed,J., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
             Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
             Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
             Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
             Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
             Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
             Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
             Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
             Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
             Weinstock,G. and Gibbs,R.A.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 238103)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3  (bases 1 to 238103)
  AUTHORS    Rat Genome Sequencing Consortium.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On May 10, 2003 this sequence version replaced gi:23264508.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             and whole genome shotgun sequencing reads assembled as described
             (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
```

```
FEATURES             Location/Qualifiers
     source          1..221586
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-170K13"
                     /clone_lib="DanioKey"
     misc_feature    1..102058
                     /note="assembly_fragment:02700
     misc_feature    102159..113284
                     /note="assembly_fragment:00743
     misc_feature    113385..121021
                     /note="assembly_fragment:00121
     misc_feature    121122..156867
                     /note="assembly_fragment:00225
     misc_feature    156968..162929
                     /note="assembly_fragment:00123
     misc_feature    163030..184102
                     /note="assembly_fragment:01032
     misc_feature    184203..221586
                     /note="assembly_fragment:01392
ORIGIN
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1   102058: contig of 102058 bp in length
* 102058: gap of 100 bp
* 102159: contig of 11126 bp in length
* 113284: gap of 100 bp
* 113285: contig of 100 bp in length
* 113384: gap of 100 bp
* 113385: contig of 7637 bp in length
* 121021: gap of 100 bp
* 121022: contig of 100 bp in length
* 121121: gap of 100 bp
* 121122: contig of 35746 bp in length
* 156867: gap of 100 bp
* 156868: contig of 100 bp in length
* 156967: gap of 100 bp
* 156968: contig of 5962 bp in length
* 162929: gap of 100 bp
* 162930: contig of 100 bp in length
* 163029: gap of 100 bp
* 163030: contig of 21073 bp in length
* 184102: gap of 100 bp
* 184103: contig of 100 bp in length
* 184202: gap of 100 bp
* 184203: contig of 37384 bp in length

```
US-10-617-443B-2_COPY_1_9 (1-9) x BX957332 (1-221586)

DB:
Query Match:
Best Local Similarity:    88.64%
Percent Similarity:      100.00%
Score:                   100.00%
Pred. No.:                39.00
Alignment Scores:         1.96e+04

Qy        2 HisLeuAlaLeuProAlaSerAla  9
            |||||||||||||||||||||||||
Db   167230 CATCTTGCCCTGCCGGCTTCAGCC 167253

                                    Length:       221586
                                    Matches:           8
                                    Conservative:      0
                                    Mismatches:        0
                                    Indels:            0
                                    Gaps:              0
```

```
Qy        2 HisLeuAlaLeuProAlaSerAla  9
            |||||||||||||||||||||||||
Db   172836 CATCTCGCCCTCCCTCGCTTCCGCC 172859
```

Search completed: June 29, 2005, 23:51:05
Job time : 337.461 secs

US-10-617-443B-2_COPY_1_9 (1-9) x AC099140 (1-238103)

DB:
Query Match: 2   Gaps: 0
Best Local Similarity: 88.64%   Indels: 0
Percent Similarity: 100.00%   Mismatches: 0
Score: 39.00   Conservative: 0
Pred. No.:   1   Matches: 8
Alignment Scores: 2.09e+04   Length: 238103

ORIGIN
              end_sequence:BH344409"
              clone_end:Sp6
              site:EcoRI
              /note="clone_boundary
    misc_feature   complement(232158..232850)
              end_sequence:BH344407"
              site:EcoRI
              clone_end:T7
              /note="clone_boundary
    misc_feature   1634.. 2403
              clone_end:T7"
              /note="wgs_end_extension
    misc_feature   1.. 1535
              /clone="CH230-52A14"
              /db_xref="taxon:10116".
              /mol_type="genomic DNA"
              /organism="Rattus norvegicus"
    source   1.. 238103
              Location/Qualifiers
FEATURES
              be preserved.
         *   237014 238103: contig of 1090 bp in length.
         *   237013: gap of unknown length
         *   236914 237013: contig of 1005 bp in length
         *   236913: gap of unknown length
         *   235909 236913: contig of 1005 bp in length
         *   235908: gap of unknown length
         *   234773 235908: contig of 1036 bp in length
         *   234772: gap of unknown length
         *   1   234672: contig of 234672 bp in length
    * as soon as it is available and the accession number will
    * This record will be updated with the finished sequence
    * runs of N, but the exact sizes of the gaps are unknown.
    * arbitrary. Gaps between the contigs are represented as
    * is not known and their order in this sequence record is
    * consists of 4 contigs. The true order of the pieces
    * NOTE: This is a 'working draft' sequence. It currently
    *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
    * NOTE: Estimated insert size may differ from sequence length
    ----
    Quality coverage: 14x in Q20 bases; sum-of-contigs estimation
    Estimated insert size: 241423; sum-of-contigs estimation
    Consensus quality: 233309 bases at least Q20
    Consensus quality: 232971 bases at least Q30
    Consensus quality: 232225 bases at least Q40
    Consensus quality: 232295 bases at least Q40
    Assembly program: Atlas 3.0;
    --------   Summary Statistics
    Center clone name: CH230-52A14
    Center project name: GGYK
    --------   Project Information
    Contact: hgsc-help@bcm.tmc.edu
    Web site: http://www.hgsc.bcm.tmc.edu/
    Center code: BCM
    Center: Baylor College of Medicine
    --------   Genome Center
    ----
    table.
    shotgun sequence only contigs will be indicated in the feature
    genome shotgun sequence reads. Both end sequences and whole genome
    contigs within a contig-scaffold that consist entirely of whole
    may extend beyond the ends of the clone and there may be sequence
    by sized gaps filled with Ns to to the estimated size. The sequence
    individual sequence contigs are ordered and oriented, and separated
    assembly (a 'contig-scaffold'). Within each contig-scaffold,
    in the feature table below represents a scaffold in the Atlas
```

OM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2005 Compugen Ltd.
GenCore version 5.1.6

Run on:    June 29, 2005, 18:37:40 ; Search time 25.2682 Seconds
           2108.485 Million cell updates/sec
           (without alignments)

Title:         US-10-617-443B-2_COPY_1_9
Perfect score: 44
Sequence:      1 MHLALPASA 9

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0  , Fgapext 7.0
               Delop  6.0  , Delext  7.0

Searched:      4390206 seqs, 295970667 residues

Total number of hits satisfying chosen parameters:    8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10617443/runat_29062005_135931_26052/app_query.fasta_1.718
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=0
-USER=US10617443 @CGN 1 1 586 @runat 29062005 135931 26052 -NCPU=6
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
           1: geneseqn1980s:*
           2: geneseqn1990s:*
           3: geneseqn1990s:*
           4: geneseqn2000s:*
           5: geneseqn2001as:*
           6: geneseqn2001bs:*
           7: geneseqn2002as:*
           8: geneseqn2002bs:*
           9: geneseqn2003as:*
          10: geneseqn2003cs:*
          11: geneseqn2003ds:*
          12: geneseqn2004as:*
          13: geneseqn2004bs:*

| c | No. | Score | | | ID | Description |
|---|---|---|---|---|---|---|
| c | 6 | 86.4 | 60940 | 9 | ADA02582 | Ada02582 Human REL |
| c | 7 | 86.4 | 60940 | 10 | ADB72320 | Adb72320 Human REL |
| c | 8 | 86.4 | 60940 | 10 | ADE95830 | Ade95830 Human REL |
| c | 9 | 86.4 | 110000 | 4 | | Continuation (20 o |
| c | 10 | 84.1 | 2196 | 2 | AAQ67901 | Aaq67901 Syndecan |
| c | 11 | 84.1 | 2196 | 2 | AAV41533 | Aav41533 Nucleotid |
| c | 12 | 84.1 | 2196 | 2 | AAV15947 | Aav15947 Mouse syn |
| c | 13 | 84.1 | 2196 | 8 | ACA60751 | Aca60751 Mouse syn |
| c | 14 | 84.1 | 21562 | 11 | ACN45004 | Acn45004 Mouse gen |
| c | 15 | 84.1 | 247544 | 12 | AAK67239 | Aak67239 Human can |
| c | 16 | 84.1 | 34 | 12 | ADQ59419 | Adq59419 Human can |
| c | 17 | 81.8 | 34 | 12 | ADJ11359 | Adj11359 PCR prime |
| c | 18 | 81.8 | 706 | 9 | ACL24588 | Acl24588 DNA clone |
| c | 19 | 81.8 | 756 | 8 | ACA22808 | Aca22808 Prokaryot |
| c | 20 | 81.8 | 757 | 8 | ACA21959 | Aca21959 Prokaryot |
| c | 21 | 81.8 | 795 | 8 | ADN62510 | Adn62510 A. thalia |
| c | 22 | 81.8 | 1077 | 3 | AAC40009 | Aac40009 Arabidops |
| c | 23 | 81.8 | 1387 | 8 | ACA57488 | Aca57488 Human adi |
| c | 24 | 81.8 | 1395 | 13 | ADQ85511 | Adq85511 Human tum |
| c | 25 | 81.8 | 1797 | 12 | ADQ64218 | Adq64218 Novel hum |
| c | 26 | 81.8 | 1854 | 4 | AAI59099 | Aai59099 Human pol |
| c | 27 | 81.8 | 1854 | 5 | AAI60885 | Aai60885 Human pol |
| c | 28 | 81.8 | 1854 | 5 | ADQ99322 | Adq99322 DNA encod |
| c | 29 | 81.8 | 1854 | 9 | ADB49082 | Adb49082 Novel hum |
| c | 30 | 81.8 | 1939 | 5 | ABA01852 | Aba01852 Novel reg |
| c | 31 | 81.8 | 2057 | 10 | ADI121987 | Adi121987 Novel hum |
| c | 32 | 81.8 | 2160 | 9 | ADA27231 | Ada27231 Human NOV |
| c | 33 | 81.8 | 2203 | 10 | ADE95581 | Ade95581 Human NOV |
| c | 34 | 81.8 | 2250 | 12 | ADO22235 | Ado22235 Human WNT |
| c | 35 | 81.8 | 2359 | 10 | ADE58551 | Ade58551 Human gen |
| c | 36 | 81.8 | 2374 | 4 | AAK78587 | Aak78587 Human imm |
| c | 37 | 81.8 | 2389 | 8 | ACC48788 | Acc48788 Human Evi |
| c | 38 | 81.8 | 2389 | 8 | ACC48792 | Acc48792 Human Evi |
| c | 39 | 81.8 | 2491 | 10 | ADA53235 | Ada53235 Human NOV |
| c | 40 | 81.8 | 2569 | 10 | ADD49028 | Add49028 Human NOV |
| c | 41 | 81.8 | 2569 | 3 | AAC93383 | Aac93383 Human sec |
| c | 42 | 81.8 | 2590 | 6 | ABN85556 | Abn85556 Human lat |
| c | 43 | 81.8 | 2629 | 10 | ADE95583 | Ade95583 Human NOV |
| c | 44 | 81.8 | 2707 | 5 | AAS44925 | Aas44925 cDNA enco |
| c | 45 | 81.8 | 2714 | 10 | ADA53450 | Ada53450 Human cod |

ALIGNMENTS

RESULT 1
ADJ11355
ID   ADJ11355 standard; cDNA; 3006 BP.
XX
AC   ADJ11355;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Human PRSS11-L cDNA encoding a serine protease SeqID 1.
XX
KW   human; PRSS11-L; gene; ss; S2 serine protease; S2/HtrA;
KW   cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;
KW   apoptotic; osteopathic; antiarthritic; tranquiliser.
XX
OS   Homo sapiens.
XX
PN   US2004005659-A1.
XX
PD   08-JAN-2004.
XX
PF   03-JUL-2002; 2002US-00189099.
XX
PR   03-JUL-2002; 2002US-00189099.
XX
FH   Key             Location/Qualifiers
FT   CDS             1011..2015
FT                   /*tag= a
FT                   /product= "PRSS11-L protein"

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 44 | 100.0 | 3006 | 12 | ADJ11355 | Adj11355 Human PRS |
| 2 | 39 | 88.6 | 754 | 6 | ABS77428 | Abs77428 Prog embr |
| 3 | 39 | 88.6 | 963 | 8 | ACA25538 | Aca25538 Prokaryot |
| 4 | 39 | 88.6 | 2157 | 8 | ACA25970 | Aca25970 Prokaryot |
| 5 | 38 | 86.4 | 578 | 10 | ADG37683 | Adg37683 Aspergill |

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

PA  (DARR/)  DARROW A L.
PA  (QLJJ/)  QL J.
PA  (CHEN/)  CHEN C.
PA  (ANDR/)  ANDRADE-GORDON P.
XX
PI  Darrow AL, Ql J, Chen C, Andrade-Gordon P;
XX
DR  WPI; 2004-081723/08.
DR  P-PSDB; ADJ11356.
XX
PS  Claim 2; SEQ ID NO 1; 28pp; English.
XX
CC  New isolated S2 serine protease nucleic acids and polypeptides, useful
PT  for diagnosing and/or treating diseases with aberrant expression or
PT  activity the S2 serine protease, such as osteoarthritis, stress and
PT  apoptotic disorders.
XX
CC  This invention relates to a novel isolated nucleic acid molecule
CC  identified as PRSS11-L that encodes an S2 serine protease. Specifically,
CC  it refers to members of the S2/HtrA serine protease family, such that it
CC  plays a role in cellular physiology and apoptosis. The present invention
CC  provides agonists, antagonists, antibodies and recombinant expression
CC  vectors useful in methods of treatment, or detection and diagnosis of
CC  diseases associated with the aberrant expression or activity of the S2
CC  serine protease, PRSS11-L. Accordingly, compositions described herein can
CC  be used via gene therapy routes to treat osteoarthritis, stress and
CC  apoptotic disorders. As such, they exhibit osteopathic, antiarthritic and
CC  tranquiliser activities. This polynucleotide sequence is the human PRSS11
CC  -L cDNA sequence of the invention.
XX
SQ  Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                166
Score:                    44.00
Percent Similarity:       100.00%
Best Local Similarity:    100.00%
Query Match:              100.00%

Length:                   3006
Matches:                  9
Conservative:             0
Mismatches:               0
Indels:                   0
Gaps:                     0
DB:                       12

US-10-617-443B-2_COPY_1_9 (1-9) x ADJ11355 (1-3006)

```
Qy    1 MetHisLeuValLeuProAlaSerAla 9
          |||||||||||||||||||||||||||
Db 1011 ATGCACCTGGCCCCTTCCCGCCAGCCCA 1037
```

RESULT 2
ID ABS77428   standard; cDNA; 754 BP.
XX
AC ABS77428;
XX
DT 12-DEC-2002 (first entry)
XX
DE Frog embryonic gene sequence Q9925836.
XX
KW Frog; ss; embryonic development; developmental disorder; microarray;
KW cell differentiation.
XX
OS Xenopus laevis.
XX
XX
PN US2002081610-A1.
XX
PD 27-JUN-2002.
XX
PF 23-JUL-2001; 2001US-00910943.
PR 21-JUL-2000; 2000US-0219658P.
XX
PA (UYRQ) UNIV ROCKEFELLER.
XX
PI Hemmati-Brivanlou A, Altmann CR;

PT Nucleic acid array containing Xenopus embryonic nucleic acids is useful
PT to identify genes involved in embryonic development, to identify
PT different types of embryonic cells, and to diagnose developmental
PT disorders.
XX
DR WPI; 2002-626534/67.
XX
PS Claim 1; Page 698; 823pp; English.
XX
CC The invention relates to a nucleic acid array, where each coordinate
CC contains a single nucleic acid species having one of 770 nucleotide
CC sequences (appearing as ABS76747-ABS77516) of a Xenopus embryonic gene
CC product, or its complement or hybridisable fragment of not less than 20
CC contiguous nucleotides of one of those sequences. Also included are
CC detecting differential expression of embryonic genes, comprising: (a)
CC contacting a nucleic acid array comprising genes expressed in embryonic
CC but not mature cells with nucleic acids from sample and control cells;
CC and (b) detecting differential hybridisation of nucleic acids from the
CC sample cells relative to the control cells; and detecting defects in
CC development, comprising: (a) contacting nucleic acids from test cells
CC undergoing development with a nucleic acid array of gene products known
CC to play a fundamental role in the development process; and (b) detecting
CC a difference in expression of a fundamental gene in the sample cells
CC relative to a standard. The invention is useful to identify genes
CC involved in embryonic development and related processes such as cell
CC differentiation. This would be useful for diagnosing developmental
CC disorders and for identifying different types of embryonic cells. The
CC present sequence is one of the 770 Xenopus embryonic cDNA sequences
XX
SQ Sequence 754 BP; 169 A; 200 C; 171 G; 207 T; 0 U; 7 Other;

Alignment Scores:
Pred. No.:   346
Score: 39.00
Percent Similarity: 100.00%
Best Local Similarity: 77.78%
Query Match: 88.64%

Length: 754
Matches: 7
Conservative: 2
Mismatches: 0
Indels: 0
Gaps: 0
DB: 6

US-10-617-443B-2_COPY_1_9 (1-9) x ABS77428 (1-754)

```
Qy    1 MetHisLeuValLeuProAlaSerAla 9
          |||::|||:::|||||||||||||
Db  327 TTACATATCGCTCTGCCGGCCCTCTGCT 353
```

RESULT 3
ID ACA25538   standard; DNA; 963 BP.
XX
AC ACA25538;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #7195.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Burkholderia fungorum.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.

Left column

```
XX
PI  Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI  Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR  WPI; 2003-029926/02.
DR  P-PSDB; ABU21668.
XX
PN  New antisense nucleic acids, useful for identifying proteins or screening
PT  for homologous nucleic acids required for cellular proliferation to
PT  isolate candidate molecules for rational drug discovery programs.
XX
XX
PS  Claim 14; SEQ ID NO 13408; 1766pp; English.
XX
CC  The invention relates to an isolated nucleic acid comprising any one of
CC  the 6213 antisense sequences given in the specification where expression
CC  of the nucleic acid inhibits proliferation of a cell. Also included are:
CC  (1) a vector comprising a promoter operably linked to the nucleic acid
CC  encoding a polypeptide whose expression is inhibited by the antisense
CC  nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC  polypeptide or its fragment whose expression is inhibited by the
CC  antisense nucleic acid; (4) an antibody capable of specifically binding
CC  the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC  proliferation or the activity of a gene in an operon required for
CC  proliferation; (7) identifying a compound that influences the activity of
CC  the gene product or that has an activity against a biological pathway
CC  required for proliferation, or that inhibits cellular proliferation (8)
CC  identifying a gene required for cellular proliferation or the biological
CC  pathway in which a proliferation-required gene or its gene product lies
CC  or a gene on which the test compound that inhibits proliferation of an
CC  organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC  compound's activity; (11) a culture comprising strains in which the gene
CC  product is overexpressed or underexpressed; (12) determining the extent
CC  to which each of the strains is present in a culture or collection of
CC  strains; or (13) identifying the target of a compound that inhibits the
CC  proliferation of an organism. The antisense nucleic acids are useful for
CC  identifying proteins or screening for homologous nucleic acids required
CC  for cellular proliferation to isolate candidate molecules for rational
CC  drug discovery programs, or for screening homologous nucleic acids
CC  required for proliferation in cells other than S. aureus, S. typhimurium,
CC  K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC  prokaryotic essential genes. Note: The sequence data for this patent did
CC  not form part of the printed specification, but was obtained in
CC  electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 963 BP; 176 A; 300 C; 322 G; 165 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                    452          Length:        963
Percent Similarity:       39.00        Matches:       8
Best Local Similarity:    88.89%       Conservative:  0
Query Match:              88.89%       Mismatches:    0
                          88.64%       Indel:         1
                          8            Gaps:          0

US-10-617-443B-2_COPY_1_9 (1-9) x ACA25538 (1-963)

Qy      1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||||
Db    739 ATGCATCTGGCCACGGCCGGCCTCGGCG 765
```

(handwritten: Leu=CTG   25/27: 92.6%)

Right column

```
RESULT 4
ID  ACA25970  standard; DNA; 2157 BP.
XX
AC  ACA25970;
XX
DT  19-JUN-2003  (first entry)
XX
DB  Prokaryotic essential gene #7627.
XX
KW  Antisense; ds; prokaryotic essential gene; cell proliferation;
KW  drug design; gene.
XX
OS  Burkholderia mallei.
XX
PN  WO20277183-A2.
XX
PD  03-OCT-2002.
XX
PA  (ELIT-) ELITRA PHARM INC.
XX
PI  Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI  Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
PR  21-MAR-2002; 2002WO-US009107.
PR  21-MAR-2001; 2001US-00815242.
PR  06-SEP-2001; 2001US-00948993.
PR  25-OCT-2001; 2001US-0342923P.
PR  25-OCT-2001; 2001US-00948993.
PR  08-FEB-2002; 2002US-00072851.
PR  06-MAR-2002; 2002US-0362699P.
XX
DR  WPI; 2003-029926/02.
DR  P-PSDB; ABU22100.
XX
PN  New antisense nucleic acids, useful for identifying proteins or screening
PT  for homologous nucleic acids required for cellular proliferation to
PT  isolate candidate molecules for rational drug discovery programs.
XX
PS  Claim 14; SEQ ID NO 13840; 1766pp; English.
XX
CC  The invention relates to an isolated nucleic acid comprising any one of
CC  the 6213 antisense sequences given in the specification where expression
CC  of the nucleic acid inhibits proliferation of a cell. Also included are:
CC  (1) a vector comprising a promoter operably linked to the nucleic acid
CC  encoding a polypeptide whose expression is inhibited by the antisense
CC  nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC  polypeptide or its fragment whose expression is inhibited by the
CC  antisense nucleic acid; (4) an antibody capable of specifically binding
CC  the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC  proliferation or the activity of a gene in an operon required for
CC  proliferation; (7) identifying a compound that influences the activity of
CC  the gene product or that has an activity against a biological pathway
CC  required for proliferation, or that inhibits cellular proliferation; (8)
CC  identifying a gene required for cellular proliferation or the biological
CC  pathway in which a proliferation-required gene or its gene product lies
CC  or a gene on which the test compound that inhibits proliferation of an
CC  organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC  compound's activity; (11) a culture comprising strains in which the gene
CC  product is overexpressed or underexpressed; (12) determining the extent
CC  to which each of the strains is present in a culture or collection of
CC  strains; or (13) identifying the target of a compound that inhibits the
CC  proliferation of an organism. The antisense nucleic acids are useful for
CC  identifying proteins or screening for homologous nucleic acids required
CC  for cellular proliferation to isolate candidate molecules for rational
CC  drug discovery programs, or for screening homologous nucleic acids
CC  required for proliferation in cells other than S. aureus, S. typhimurium,
CC  K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC  prokaryotic essential genes. Note: The sequence data for this patent did
CC  not form part of the printed specification, but was obtained in
CC  electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 2157 BP; 350 A; 702 C; 761 G; 344 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                    1.08e+03     Length:        2157
Percent Similarity:       39.00        Matches:       8
Best Local Similarity:    88.89%       Conservative:  0
Query Match:              88.89%       Mismatches:    0
                          88.64%       Indel:         1
                          8            Gaps:          0

US-10-617-443B-2_COPY_1_9 (1-9) x ACA25970 (1-2157)
```

Qy

RESULT 5
ADG37683

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
         |||||||||||||||||||||||||
Db 1933 ATGCATCTGGCCACGCCGCCTCCGCG 1959

Db
ID   ADG37683 standard; DNA; 578 BP.
AC   ADG37683;
XX
DT   26-FEB-2004 (first entry)
XX
DE   Aspergillus solid-culture DNA #394.
XX
KW   ds; filamentous-fungus; Aspergillus fermentation; solid culture.
XX
OS   Aspergillus oryzae.
XX
PN   JP2003180365-A.
XX
PD   02-JUL-2003.
XX
PF   17-DEC-2001; 2001JP-0383645.
XX
PR   17-DEC-2001; 2001JP-0383645.
XX
PA   (DOKU-) DOKURITSU GYOSEI HOJIN SHURUI SOGO KENKY.
XX
DR   WPI; 2003-819301/77.
XX
PT   New DNA derived from Aspergillus, expressed by filamentous-fungi, useful for measuring Aspergillus fermentation conditions.
PS   Claim 2; SEQ ID NO 394; 302pp; Japanese.
XX
CC   The invention relates to a DNA expressed by filamentous-fungi. The DNA is useful for detecting filamentous-fungi and is useful for measuring cultured in a solid medium, useful for detecting the DNA is Aspergillus fermentation conditions. A method of detecting the DNA is useful for measuring the actual growth state of Aspergillus quantitatively and the expression level of desired gene specifically. The present sequence is used in the exemplification of the present invention.
XX
SQ   Sequence 578 BP; 133 A; 154 C; 126 G; 165 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 578
Score: 406   Matches: 7
                       38.00   Conservative: 2
Percent Similarity: 100.00%   Mismatches: 0
Best Local Similarity: 77.78%   Indels: 0
Query Match: 86.36%   Gaps: 0
                       10

RESULT 6
ADA02582/c
ID   ADA02582 standard; DNA; 60940 BP.
AC   ADA02582;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Human REL carcinoma associated gene, SEQ ID NO:1100.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast; prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening; gene; ds.
XX

OS   Homo sapiens.
XX
PN   WO200305714 6-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein, useful for preparing compositions for treating carcinomas.
PS   Claim 1; SEQ ID NO 1100; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic acid sequences from mouse and human (ADA01482-ADA03094), and to recombinant carcinoma associated proteins (CAP) encoded by them. The invention also encompasses expression vectors and host cells comprising a CA nucleic acid, a polypeptide (especially an antibody) that specifically binds to the protein, and a biochip comprising CA nucleic acid or fragments thereof. The sequences of the invention were identified using oncogenic retroviruses, which insert into the genome of the host organism at random. Many of these do not carry transduced host oncogenes or pathogenic trans-acting viral genes, meaning that cancer incidence is a direct consequence of the effects of proviral integration into host protooncogenes. The CA nucleic acid sequences can be used to diagnose carcinoma (especially breast cancer, prostate cancer, lymphoma or leukaemia) or a propensity to carcinoma by determination of the sequence of a CA gene, or by determination of CA gene expression in particular tissues. CA nucleic acids, proteins and antibodies are also useful as therapeutic agents and in screening and evaluating drug candidates. The present sequence represents a specifically claimed human CA nucleic acid sequence of the invention. Note: The complete sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 60940 BP; 16572 A; 11094 C; 10947 G; 18968 T; 0 U; 3359 Other;

Alignment Scores:
Pred. No.:   Length: 60940
Score: 6.34e+04   Matches: 8
                       38.00   Conservative: 0
Percent Similarity: 88.89%   Mismatches: 1
Best Local Similarity: 88.89%   Indels: 0
Query Match: 86.36%   Gaps: 0
                       9

US-10-617-443B-2_COPY_1_9 (1-9) x ADA02582 (1-60940)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
         ||||||||||||||||| |||||
Db 11468 ATGCATTTAGCCCTCCCTGCCTTTGCT 11442

RESULT 7
ADB72320/c
Db
ID   ADB72320 standard; DNA; 60940 BP.
AC   ADB72320;
XX
DT   04-DEC-2003 (first entry)
XX
DE   Human REL gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphoma; cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX

```
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
DR
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-0004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-0099722.
PR   20-DEC-2001; 2001US-0034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW, Engelhard EK;
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 148; 2304pp; English.
XX
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 60940 BP; 16572 A; 11094 C; 10947 G; 18968 T; 0 U; 3359 Other;

US-10-617-443B-2_COPY_1_9 (1-9) x ADB72320 (1-60940)

Qy      1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||| |||
Db  11468 ATGCATTTAGCCCTCCCTGCCTTTGCT 11442

Alignment Scores:
Pred. No.:
Score:                   6.34e+04        Length:           60940
Percent Similarity:      38.00           Matches:          8
Best Local Similarity:   88.89%          Conservative:     0
                         88.89%          Mismatches:       1
Query Match:             86.36%          Indels:           0
                         10              Gaps:             0

RESULT 8
ID   ADB95830/c
AC   ADB95830;
XX
DT   12-FEB-2004 (first entry)
XX
DE   Human REL gene genomic DNA sequence.
XX
KW   cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
KW   lymphoma; breast cancer; prostate cancer; leukaemia; de; human; REL.
XX
OS   Homo sapiens.
XX
PN   WO200303484-A2.
XX
PD   15-MAY-2003.
XX
PF   08-NOV-2002; 2002WO-US0360071.
XX
PR   08-NOV-2002; 2002WO-US052482.
PA   (SAGR-) SAGRES DISCOVERY.
```

```
XX
PI   Morris DW, Engelhard EK;
XX
DR   WPI; 2003-441462/41.
XX
PT   New carcinoma associated nucleic acids and proteins, useful for screening
PT   drug candidates, or for diagnosing and treating carcinomas, e.g.
PT   lymphoma, breast cancer, prostate cancer or leukemia.
XX
PS   Claim 1; SEQ ID NO 88; 793pp; English.
XX
XX
CC   This invention relates to novel recombinant nucleic acids for use in
CC   diagnosis and treatment of cancer, especially carcinomas, as well as the
CC   use of compositions in screening methods. The compositions of the
CC   invention may have cytostatic activity whilst the disclosed sequences may
CC   be useful for gene therapy. The carcinoma associated nucleic acids and
CC   proteins are useful for diagnosing and treating carcinomas, for example
CC   lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC   drug candidates or bioactive agents capable of binding to, or modulating
CC   the activity of, a carcinoma associated protein. The present sequence is
CC   the genomic DNA sequence of the human REL gene which is a carcinoma
CC   associated gene of the invention.
XX
SQ   Sequence 60940 BP; 16572 A; 11095 C; 10946 G; 18968 T; 0 U; 3359 Other;

US-10-617-443B-2_COPY_1_9 (1-9) x ADB95830 (1-60940)

Qy      1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||| |||
Db  11468 ATGCATTTAGCCCTCCCTGCCTTTGCT 11442

Alignment Scores:
Pred. No.:
Score:                   6.34e+04        Length:           60940
Percent Similarity:      38.00           Matches:          8
Best Local Similarity:   88.89%          Conservative:     0
                         88.89%          Mismatches:       1
Query Match:             86.36%          Indels:           0
                         10              Gaps:             0

RESULT 9
AAI99682_19/c
Continuation (20 of 45) of AAI99682 from base 1900001 (Mycobacterium tuberculosis strain
Sequence split into 45 fragments              LOCUS AAI99682 Accession Aai99682
```

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000

```
WP  AAI99682_29    2900001    3010000
WP  AAI99682_30    3000001    3110000
WP  AAI99682_31    3100001    3210000
WP  AAI99682_32    3200001    3310000
WP  AAI99682_33    3300001    3410000
WP  AAI99682_34    3400001    3510000
WP  AAI99682_35    3500001    3610000
WP  AAI99682_36    3600001    3710000
WP  AAI99682_37    3700001    3810000
WP  AAI99682_38    3800001    3910000
WP  AAI99682_39    3900001    4010000
WP  AAI99682_40    4000001    4110000
WP  AAI99682_41    4100001    4210000
WP  AAI99682_42    4200001    4310000
WP  AAI99682_43    4300001    4410000
WP  AAI99682_44    4400001    4411529
```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:

```
1.19e+05
38.00
100.00%
77.78%
86.36%
```

Length: 110000
Matches: 7
Conservative: 2
Mismatches: 0
Indels: 0
Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9

US-10-617-443B-2_COPY_1_9 (1-9) x AAI99682_19 (1-110000)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
          :::||||||||:|:||||||||||||||
Db 87705 GTTCACTTAGCGATGCCGGCGGTCAGCT 87679

ID AAQ67901/c
AC AAQ67901;
XX
XX
DT 25-MAR-2003 (revised)
DT 08-DEC-1994 (first entry)
XX
DE Syndecan gene enhancer element.
XX
KW Syndecan; enhancer; differentiation; tumor; therapy; hair growth; ss.
XX
OS Mus sp.
XX
PN WO9412162-A1.
XX
PD 09-JUN-1994.
XX
PF 01-DEC-1993; 92US-00988427.
XX
PR 01-DEC-1992; 93WO-FI000514.
XX
PI Leppae SM, Mali MS, Vihinen TA;
PA (VIHI/) VIHINEN T A.
PA (MALI/) MALI M S.
PA (LEPP/) LEPPAE S M.
PA (JALK/) JALKANEN M T.
PA (JAAK/) JAAKKOLA P M.
PA (AUVI/) AUVINEN P O V.
PA (ALAN/) ALANEN-KURKI L M.
PA (WAER/) WAERRI A M.
XX
XX
PI Waerri AM, Alanen-Kurki LM, Auvinen POV, Jaakkola PM, Jalkanen MT;
```
Db  87705 GTTCACTTAGCGATGCCGGCGGTCAGCT 87679

RESULT 10
AAQ67901/c
AAQ67901;
standard; DNA; 2196 BP.

WPI; 1994-199926/24.

Disclosure; Page 40-41; 65pp; English.

Syndecan stimulation of cellular differentiation - useful for decreasing tumour growth used to promote hair growth.

---

```
CC  The mouse syndecan gene enhancer, located 8-10 kb upstream from the
CC  initiation site, is given in AAQ67901. Manipulation of the enhancer can
CC  be used either to slow or prevent tumor growth or to promote
CC  differentiation of specific cell types, e.g. epidermal cells to promote
CC  hair formation. The complete mouse syndecan gene and its encoded protein
CC  are given in AAQ67902 and AAR55276. (Updated on 25-MAR-2003 to correct PN
CC  field.) (Updated on AAQ67902 and AAR55276. (Updated on 25-MAR-2003 to
CC  field.) (Updated on 25-MAR-2003 to correct PI field.)
CC  2003 to correct PI field.)
```

SQ Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;

US-10-617-443B-2_COPY_1_9 (1-9) x AAQ67901 (1-2196)

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:

```
2.7e+03
37.00
100.00%
87.50%
84.09%
```

Length: 2196
Matches: 7
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSer 8
          |||:|::||||||||||||||||||
Db   1143 ATGCATGTGGGCACTCCCTGCCTCA 1120

(handwritten: Len=CTG→26/27 = 96.3%)

RESULT 11
ID AAV41533/c
AC AAV41533;
XX
XX
DT 28-SEP-1998 (first entry)
XX
DE Nucleotide sequence of mouse syndecan gene.
XX
KW Mouse; syndecan gene; syndecan enhancer element; expression vector;
KW wound healing; ss.
XX
OS Mus sp.
XX
PN WO9824921-A1.
XX
PD 11-JUN-1998.
XX
PF 02-DEC-1997; 97WO-FI000748.
XX
PR 02-DEC-1996; 96US-00760534.
XX
PA (BIOT-) BIOTIE THERAPIES LTD OY.
XX
PI Jalkanen M, Jaakkola P, Vihinen T;
XX

WPI; 1998-333338/29.

New syndecan enhancer element - useful in expression vector(s) for promoting wound healing.

Disclosure; Page 29-30; 57pp; English.

This is the nucleotide sequence of the mouse syndecan gene, used in the method of the invention involving the novel syndecan enhancer element, useful in expression vectors. The vector can be used to produce a host cell for expressing a structural gene. The expression vector can be used to differentially express a gene at a wound site, and promote wound healing

SQ Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:

```
2.7e+03
37.00
100.00%
87.50%
```

Length: 2196
Matches: 7
Conservative: 1
Mismatches: 0

RESULT 12

```
Qy        Query Match:        84.09%        Indels:    0
DB:                           2             Gaps:      0

US-10-617-443B-2_COPY_1_9 (1-9) x AAV41533 (1-2196)

Qy      1 MetHisLeuAlaLeuProAlaSer 8
          ||||::||
Db   1143 ATGCATGTGGCACTCCCTGCCTCA 1120
```

(handwritten: Leu = CTG ; 26/27 = 96.3%)

```
ID   AAV15947 standard; DNA; 2196 BP.
XX
AC   AAV15947;
XX
DT   28-MAY-1998  (first entry)
XX
DE   Mouse syndecan gene enhancer element.
XX
KW   Syndecan; tumour suppression; tissue regeneration; enhancement; mouse;
KW   wound healing; enhancer element; ds.
XX
OS   Mus. sp.
XX
PN   US5726058-A.
XX
PD   10-MAR-1998.
XX
PF   07-JUN-1995;   95US-00472217.
PR   01-DEC-1992;   92US-00988427.
PR   01-DEC-1993;   93WO-PI000514.
PR   07-MAR-1994;   94US-00206186.
XX
PA   (AUVI/) AUVINEN P.
PA   (MALI/) MALI M.
PA   (VIHI/) VIHINEN T.
PA   (WAER/) WAERRI A.
PA   (JALK/) JALKANEN M.
PA   (ALAN/) ALANEN-KURKI L.
PA   (LEPP/) LEPPAE S.
PA   (JAAK/) JAAKKOLA P.
XX
PI   Auvinen P, Jalkanen M, Vihinen T;
PI   Jaakkola P, Waerri A, Leppae S, Mali M, Alanen-Kurki L;
XX
DR   WPI; 1998-192770/17.
XX
PT   New mouse syndecan gene sequences - useful for, e.g. suppressing tumour
PT   growth or promoting tissue regeneration in processes such as wound
PT   healing.
XX
PS   Claim 1; Fig 4; 48pp; English.
XX
CC   This is the mouse syndecan gene enhancer element sequence located
CC   approximately 9 kb upstream from the transcription initiation site. A 350
CC   base pair fragment (AAV15948) of this purified enhancing DNA molecule
CC   enhances the expression of a gene operably linked to the promoter of the
CC   mouse syndecan gene in 3T3 cells following treatment with TGF- beta and
CC   bFGF when the fragment is operably linked to the promoter. A purified DNA
CC   molecule comprising a portion of the nucleotide residues 3538-3888 of the
CC   mouse syndecan genomic sequence (AAV15946) suppresses expression of a
CC   gene operably linked to the promoter of the mouse syndecan gene in S115
CC   cells treated with testosterone. Host cells can be transfected with
CC   vectors which contain either the enhancing or suppressing DNA molecules.
CC   The products may be used to alter the differentiated state of a host cell
CC   by altering its expression of syndecan, e.g. to induce and regulate
CC   syndecan expression, especially in cells which exhibit a malignant
CC   phenotype, regardless of the origin of transformation. The products can
CC   be used to produce therapeutics for suppressing tumour growth. They may
CC   enhance the syndecan expression in a host cell, by enhancing its gene
CC   transcription, especially in malignant or normal cells, and therefore
CC   promote tissue regeneration, especially in processes such as wound
CC   healing
XX
SQ   Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;
```

RESULT 13

```
Alignment Scores:
Pred. No.:              2.7e+03       Length:         2196
Score:                  37.00        Matches:           7
Percent Similarity:     100.00%      Conservative:      1
Best Local Similarity:  87.50%       Mismatches:        0
Query Match:            84.09%       Indels:            0
                        2            Gaps:              0

US-10-617-443B-2_COPY_1_9 (1-9) x AAV15947 (1-2196)

Qy      1 MetHisLeuAlaLeuProAlaSer 8
          |||||::|||||||||||||||||
Db   1143 ATGCATGTGGCACTCCCTGCCTCA 1120

ACA60751/c
ID   ACA60751 standard; DNA; 2196 BP.
XX
AC   ACA60751;
XX
DT   27-JUN-2003  (first entry)
XX
DE   Mouse syndecan enhancer element, FGF-inducible Response Element (FiRE).
XX
KW   Mouse; ds; syndecan enhancer element; tissue regeneration;
KW   syndecan regulation; syndecan expression; cutaneous wound healing; FiRE;
KW   hair growth; FGF-inducible Response Element.
XX
OS   Mus. sp.
XX
FH   Key            Location/Qualifiers
FH
FT   enhancer       1150..1499
FT                  /tag= a
FT                  /bound_moiety= "FGF-2 dependent nuclear factor"
FT                  /note= "DNA binding motif 5"
FT   misc_binding   1450..1468
FT                  /tag= d
FT                  /bound_moiety= "FGF-2 dependent nuclear factor"
FT                  /note= "DNA binding motif 4"
FT   misc_binding   1477..1491
FT                  /tag= e
FT                  /bound_moiety= "FGF-2 dependent nuclear factor"
FT                  /note= "DNA binding motif 3"
FT   misc_binding   1497..1511
FT                  /tag= f
FT                  /bound_moiety= "FGF-2 independent nuclear factor"
FT                  /note= "DNA binding motif 2"
FT   misc_binding   1526..1551
FT                  /tag= g
FT                  /bound_moiety= "FGF-2 independent nuclear factor"
FT                  /note= "DNA binding motif 1"
FT   enhancer       1331..1614
FT                  /tag= b
FT                  /standard_name= "Enhancer"
FT   misc_binding   1407..1446
FT                  /tag= c
FT                  /standard_name= "Enhancer"
FT                  /note= "Specifically claimed in claim 1"
XX
PN   US6492344-B1.
XX
PD   10-DEC-2002.
XX
PF   21-JUN-1999;   99US-00336757.
PR   01-DEC-1993;   93WO-PI000514.
PR   07-MAR-1994;   94US-00206186.
XX
PA   (BIOT-) BIOTIE THERAPIES CORP.
```

Left column

```
Alignment Scores:
Pred. No.:              2.7e+03      Length:        2196
Score:                  37.00        Matches:       7
Percent Similarity:     100.00%      Conservative:  1
Percent Identity:       87.50%       Mismatches:    0
Best Local Similarity:  84.09%       Indels:        0
Query Match:            8            Gaps:          0
```

US-10-617-443B-2_COPY_1_9 (1-9) x ACA60751 (1-2196)

```
Qy    1 MetHisLeuAlaLeuProAlaSer 8
        ||||||:|||||||||||||||||
Db 1143 ATGCATGTGGCACTCCCTGCCTCA 1120
```

```
XX
SQ   Sequence 2196 BP; 509 A; 582 C; 536 G; 569 T; 0 U; 0 Other;
XX
CC   The invention relates to a method of promoting cutaneous wound healing
CC   which comprises directly introducing to a cutaneous wound site a
CC   recombinant expression vector comprising the syndecan enhancer element,
CC   FGF-inducible Response Element (FiRE), operably linked to a FiRE-
CC   activated promoter which is operably linked to a structural gene encoding
CC   a growth factor that promotes the cutaneous wound healing. The method is
CC   for promoting regeneration of tissue e.g. promotion of hair growth. The
CC   invention enhances syndecan expression via administration of growth
CC   factors. The enhancer element regulates the expression of syndecan. The
CC   present sequence represents the mouse syndecan enhancer element, FGF-
CC   inducible Response Element (FiRE).
XX
PS   Example 8; Fig 4; 84pp; English.
XX
PT   Promotion of cutaneous wound healing comprises introducing to wound site
PT   recombinant expression vector comprising syndecan enhancer element linked
PT   to promoter which is linked to structural gene encoding growth factor.
XX
PI   Jalkanen M, Jaakkola P, Vihinen T;
XX
DR   WPI; 2003-391071/37.
XX

RESULT 14
ACN45004
ID   ACN45004 standard; DNA; 21562 BP.
AC   ACN45004
XX
DT   18-NOV-2004 (first entry)
XX
DE   Mouse genomic sequence mCG6875.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX
OS   Mus musculus.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 1735; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
```

Right column

```
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the effect of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 21562 BP; 5504 A; 4710 C; 4598 G; 5450 T; 0 U; 1300 Other;
```

US-10-617-443B-2_COPY_1_9 (1-9) x ACN45004 (1-21562)

```
Qy     1 MetHisLeuAlaLeuProAlaSerAla 9
         :||||||||||||||||::|||||
Db 11652 ATCCATCTTGCCTGCCATCCAGTGCA 11678
```

```
Alignment Scores:
Pred. No.:              3.21e+04     Length:        21562
Score:                  37.00        Matches:       7
Percent Similarity:     100.00%      Conservative:  2
Percent Identity:       77.78%       Mismatches:    0
Best Local Similarity:  84.09%       Indels:        0
Query Match:            11           Gaps:          0
```

```
RESULT 15
AAK67239/c
ID   AAK67239 standard; DNA; 30393 BP.
AC   AAK67239;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0198874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   14-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   14-AUG-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
```

PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0262279P.
PR   22-AUG-2000;   2000US-0262279P.
PR   22-AUG-2000;   2000US-0262681P.
PR   22-AUG-2000;   2000US-0226668P.
PR   22-AUG-2000;   2000US-0271182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   01-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   21-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   14-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   13-OCT-2000;   2000US-0236370P.
PR   13-OCT-2000;   2000US-0239935P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0270040P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0209060P.
PR   20-OCT-2000;   2000US-0241826P.
PR   13-OCT-2000;   2000US-0239377P.
PR   01-NOV-2000;   2000US-0246617P.
PR   20-OCT-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0264476P.
PR   08-NOV-2000;   2000US-0264477P.
PR   08-NOV-2000;   2000US-0264478P.
PR   08-NOV-2000;   2000US-0244523P.
PR   08-NOV-2000;   2000US-0244524P.

PR   08-NOV-2000;   2000US-0246625P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   08-NOV-2000;   2000US-0249207P.
PR   08-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   01-DEC-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   06-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2000US-0259678P.

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Baraah SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (II) by expressing inactive proteins or to
CC   supplement the patients own production of (II). Additionally, (II)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   Length: 30393
Score: 4.66e+04   Matches: 7
Percent Similarity: 37.00   Conservative: 1
Best Local Similarity: 88.89%   Mismatches: 1
Query Match: 77.78%   Indels: 0
DB: 84.09%   Gaps: 0
                       4

US-10-617-443B-2_COPY_1_9 (1-9) x AAK67239 (1-30393)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
            |||||||||| |||||| :||||||||
Db   16409 ATGCACCTCACCCTCCCCTAGTAGCGCT 16383

Search completed: June 29, 2005, 21:40:29
Job time : 50.2682 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 29, 2005, 21:23:47 ; Search time 28.1808 Seconds
(without alignments)
1996.208 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Title: US-10-617-443B-2_COPY_1_9
Perfect score: 44
Sequence: 1 MHLALPASA 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6067389 seqs, 3125258755 residues

Total number of hits satisfying chosen parameters: 12134778

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US10617443/runat_29062005_135933_26126/app_query.fasta_1.718
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000000 -USER=US10617443 @CGN_1_1_582 @runat_29062005_135933_26126
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10J_NEW_PUB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%	Query Match Length	DB	ID	Description
1	44	100.0	3006	17	US-10-189-099A-1	Sequence 1, Appli
2	44	100.0	3006	21	US-10-617-443B-1	Sequence 1, Appli
3	39	88.6	754	9	US-09-910-943-654	Sequence 654, App
4	39	88.6	963	17	US-10-282-122A-13408	Sequence 13408, A
5	39	88.6	2157	17	US-10-282-122A-13840	Sequence 13840, A
6	38	86.4	1500	19	US-10-437-963-43712	Sequence 43712, A
7	38	86.4	1581	18	US-10-424-599-10035	Sequence 10035, A
8	38	86.4	60940	18	US-10-052-482-88	Sequence 88, Appl
9	37	84.1	582	13	US-10-027-632-270357	Sequence 270357,
10	37	84.1	582	17	US-10-027-632-270357	Sequence 270357,
11	37	84.1	4513	19	US-10-437-963-44767	Sequence 44767, A
12	37	84.1	21562	13	US-10-087-192-1735	Sequence 1735, Ap
13	37	84.1	247544	19	US-10-322-696-55	Sequence 55, Appl
14	36	81.8	34	17	US-10-189-099A-5	Sequence 5, Appli
15	36	81.8	34	21	US-10-617-443B-5	Sequence 5, Appli
16	36	81.8	281	20	US-10-425-115-88970	Sequence 88970, A
17	36	81.8	557	19	US-10-437-963-16475	Sequence 16475, A
18	36	81.8	582	13	US-10-027-632-15114	Sequence 15114, A
19	36	81.8	582	13	US-10-027-632-15115	Sequence 15115, A
20	36	81.8	582	13	US-10-027-632-15116	Sequence 15116, A
21	36	81.8	582	17	US-10-027-632-15114	Sequence 15114, A
22	36	81.8	582	17	US-10-027-632-15115	Sequence 15115, A
23	36	81.8	582	17	US-10-027-632-15116	Sequence 15116, A
24	36	81.8	756	17	US-10-282-122A-10678	Sequence 10678, A
25	36	81.8	757	17	US-10-282-122A-9829	Sequence 9829, Ap
26	36	81.8	778	20	US-10-425-115-166607	Sequence 166607,
27	36	81.8	953	13	US-10-027-632-121792	Sequence 121792,
28	36	81.8	953	13	US-10-027-632-121792	Sequence 121792,
29	36	81.8	1138	13	US-10-027-632-118836	Sequence 118836,
30	36	81.8	1138	17	US-10-027-632-118836	Sequence 118836,
31	36	81.8	1854	15	US-10-037-270-992	Sequence 992, App
32	36	81.8	1854	17	US-10-117-722-992	Sequence 992, App
33	36	81.8	1922	20	US-10-425-115-117912	Sequence 117912,
34	36	81.8	1967	19	US-10-437-963-87894	Sequence 87894, A
35	36	81.8	2160	17	US-10-364-888-7	Sequence 7, Appli
36	36	81.8	2203	17	US-10-309-290-113	Sequence 113, App
37	36	81.8	2250	16	US-10-285-976-22	Sequence 22, Appl
38	36	81.8	2250	21	US-10-847-972-21	Sequence 21, Appl
39	36	81.8	2491	17	US-10-094-749-803	Sequence 803, App
40	36	81.8	2566	18	US-10-336-003A-1	Sequence 1, Appli
41	36	81.8	2573	19	US-10-767-701-14417	Sequence 14417, A
42	36	81.8	2629	17	US-10-291-172-6	Sequence 6, Appli
43	36	81.8	2707	17	US-10-309-290-115	Sequence 115, App
44	36	81.8	2707	18	US-10-221-278-6	Sequence 6, Appli
45	36	81.8	2714	17	US-10-094-749-1018	Sequence 1018, Ap

RESULT 1
US-10-189-099A-1
; Sequence 1, Application US/10189099A
; Publication No. US20040005659A1
; Publication No. US2004005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

ALIGNMENTS

Pred. No. is the number of results predicted by chance to have a

;   LENGTH: 3006
;   TYPE: CDNA
;   ORGANISM: Homo sapiens
US-10-189-099A-1

Alignment Scores:
Pred. No.:
Score: 54.2   Length: 3006
Percent Similarity: 44.00   Matches: 9
Best Local Similarity: 100.00%   Conservative: 9
Query Match: 100.00%   Mismatches: 0
                        100.00%   Indels: 0
                        17   Gaps: 0
DB:

Qy   US-10-617-443B-2_COPY_1_9 (1-9) x US-10-189-099A-1 (1-3006)

      1 MetHisLeuAlaLeuProAlaSerAla 9
        |||||||||||||||||||||||||||
Db   1011 ATGCACCTGGCCCTTCCCGCCAGCGCA 1037

RESULT 2
US-10-617-443B-1
;   Sequence 1, Application US/10617443B
;   Publication No. US2005001977A1
;   GENERAL INFORMATION:
;   APPLICANT: Darrow, Andrew L
;   APPLICANT: Qi, Jian-shen
;   APPLICANT: Chen, Cailin
;   APPLICANT: Andrade-Gordon, Patricia
;   TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
;   FILE REFERENCE: ORT-1644
;   CURRENT APPLICATION NUMBER: US/10/617,443B
;   CURRENT FILING DATE: 2003-07-02
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: Patentin Ver. 2.1
;   SEQ ID NO 1
;   LENGTH: 3006
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-617-443B-1

Alignment Scores:
Pred. No.:
Score: 54.2   Length: 3006
Percent Similarity: 44.00   Matches: 9
Best Local Similarity: 100.00%   Conservative: 9
Query Match: 100.00%   Mismatches: 0
                        100.00%   Indels: 0
                        21   Gaps: 0
DB:

Qy   US-10-617-443B-2_COPY_1_9 (1-9) x US-10-617-443B-1 (1-3006)

      1 MetHisLeuAlaLeuProAlaSerAla 9
        |||||||||||||||||||||||||||
Db   1011 ATGCACCTGGCCCTTCCCGCCAGCGCA 1037

RESULT 3
US-09-910-943-654
;   Sequence 654, Application US/0991094J
;   Patent No. US2002081610A1
;   GENERAL INFORMATION:
;   APPLICANT: Hemmati-Brivanlou, Ali
;   APPLICANT: Altman, Curtis
;   TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
;   FILE REFERENCE: 7529/IG148US1
;   CURRENT APPLICATION NUMBER: US/09/910,943
;   CURRENT FILING DATE: 2001-07-23
;   NUMBER OF SEQ ID NOS: 742
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 654
;   LENGTH: 754
;   TYPE: DNA
;   ORGANISM: Xenopus laevis
;   FEATURE:
;   NAME/KEY: misc_feature

;   LOCATION: (1)..(754)
;   OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-654

Alignment Scores:
Pred. No.:
Score: 147   Length: 754
Percent Similarity: 39.00   Matches: 7
Best Local Similarity: 100.00%   Conservative: 2
Query Match: 77.78%   Mismatches: 0
                        88.64%   Indels: 0
                        9   Gaps: 0
DB:

Qy   US-10-617-443B-2_COPY_1_9 (1-9) x US-09-910-943-654 (1-754)

      1 MetHisLeuAlaLeuProAlaSerAla 9
        :::||||:::||||||||||||||||||
Db   327 TTACATATCGCTCTTGCCCGGCCTCTGCT 353

RESULT 4
US-10-282-122A-13408
;   Sequence 13408, Application US/10282122A
;   Publication No. US2004029129A1
;   GENERAL INFORMATION:
;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John
;   APPLICANT: Carr, Grant
;   APPLICANT: Yamamoto, Robert
;   APPLICANT: Forsyth, R.
;   APPLICANT: Xu, H.
;   TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;   FILE REFERENCE: ELITRA_034A
;   CURRENT APPLICATION NUMBER: US/10/282,122A
;   CURRENT FILING DATE: 2003-02-20
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/230,335
;   PRIOR FILING DATE: 2000-09-06
;   PRIOR APPLICATION NUMBER: 60/230,347
;   PRIOR FILING DATE: 2000-09-09
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/267,636
;   PRIOR FILING DATE: 2001-02-09
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 78614
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 13408
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Burkholderia fungorum
US-10-282-122A-13408

Alignment Scores:
Pred. No.:
Score: 189   Length: 963
Percent Similarity: 39.00   Matches: 8
Percent Similarity: 88.89%   Conservative: 0
Best Local Similarity: 88.89%   Mismatches: 1

Query Match: 88.64%
DB: 17

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-282-122A-13408 (1-963)

```
Qy    1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||||||
Db  739 ATGGCATCTCGCCACGCCGCCGGCTCCGCG 765
```

RESULT 5
US-10-282-122A-13840
; Sequence 13840, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13840
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Burkholderia mallei
US-10-282-122A-13840

Alignment Scores:
Pred. No.:
Score: 429   Length: 2157
Percent Similarity: 39.00   Matches: 8
Best Local Similarity: 88.89%   Conservative: 8
Query Match: 88.64%   Mismatches: 1
DB: 17   Indels: 0
                              Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-282-122A-13840 (1-2157)

```
Qy    1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||||||||||||||||
Db 1933 ATGCATCTCGCCACGCCGCCGCTCCCG 1959
```

RESULT 6
US-10-437-963-43712
; Sequence 43712, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 43712
; LENGTH: 1500
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_46842C.1
US-10-437-963-43712

Alignment Scores:
Pred. No.:
Score: 479   Length: 1500
Percent Similarity: 38.00   Matches: 7
Best Local Similarity: 100.00%   Conservative: 2
Query Match: 77.78%   Mismatches: 0
DB: 19   Indels: 0
                              Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-437-963-43712 (1-1500)

```
Qy    1 MetHisLeuAlaLeuProAlaSerAla 9
          |||||||||||||:::||
Db  201 CTCCACCTCGGCTCTCCCCGCGGGCCGCA 227
```

RESULT 7
US-10-424-599-10035
; Sequence 10035, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 10035
; LENGTH: 1581
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109071C.1
US-10-424-599-10035

Alignment Scores:
Pred. No.:
Score: 506   Length: 1581
Percent Similarity: 38.00   Matches: 7
Best Local Similarity: 100.00%   Conservative: 2
Query Match: 86.36%   Mismatches: 0
DB: 18   Indels: 0
                              Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-424-599-10035 (1-1581)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
              ::|||||||||||||||||||||:::||
Db   1155 CTGCAtCTAGCGCCTCCCAGCAGCTGCA 1181

RESULT 8
US-10-052-482-88/c
; Sequence 88, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 88
;   LENGTH: 60940
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5047)..(7943)
;   OTHER INFORMATION: "n" at positions 5047 to 7943 can be any base
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (35382)..(35843)
;   OTHER INFORMATION: "n" at positions 35382 to 35843 can be any base
US-10-052-482-88

Alignment Scores:
Pred. No.:
Score: 2.09e+04
Percent Similarity: 38.00
Best Local Similarity: 88.89%
Query Match: 86.36%
                            18

Length: 60940
Matches: 8
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
              |||||||||||||||||||||||| |||
Db   11468 ATGCAtTTAGCCCTCCCTGCCCTTTGCT 11442

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-052-482-88 (1-60940)

RESULT 9
US-10-027-632-270357/c
; Sequence 270357, Application US/10027632
; Publication No. US20019837iA1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 270357
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-270357
;   LENGTH: 582

Alignment Scores:
Pred. No.:
Score: 295
Percent Similarity: 37.00
Best Local Similarity: 88.89%
Query Match: 77.78%
                            13

Length: 582
Matches: 7
Conservative: 1
Mismatches: 1
Indels: 0
Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
              ::||||||| |||||||||||||||
Db   495 CTACACCTGGTACTGCCTGCSTCCGCT 469

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-027-632-270357 (1-582)

RESULT 10
US-10-027-632-270357/c
; Sequence 270357, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 270357
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-270357
;   LENGTH: 582

Alignment Scores:
Pred. No.:
Score: 295
Percent Similarity: 37.00
Best Local Similarity: 88.89%
Query Match: 77.78%
                            17

Length: 582
Matches: 7
Conservative: 1
Mismatches: 1
Indels: 0
Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
              ::||||||| |||||||||||||||
Db   495 CTACACCTGGTACTGCCTGCSTCCGCT 469

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-027-632-270357 (1-582)

RESULT 11
US-10-437-963-44767

; Sequence 44767, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 44767
; LENGTH: 4513
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47796C1.1
US-10-437-963-44767

Alignment Scores:
Pred. No.:
Score: 2.38e+03   Length: 4513
Percent Similarity: 37.00   Matches: 7
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 87.50%   Mismatches: 0
DB: 19   Indels: 0
    84.09%   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-437-963-44767 (1-4513)

Qy   2 HisLeuAlaLeuProAlaSerAla 9
    ||||||||||:|||||||||||||
Db   4231 CACCTAGCCATGCCTGCCTCCGCG 4254

RESULT 12
US-10-087-192-1735
; Sequence 1735, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1735
; LENGTH: 21562
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(21562)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1735

Alignment Scores:
Pred. No.:   1.17e+04
Score: 37.00   Length: 21562
Percent Similarity: 100.00%   Matches: 7
Percent Similarity: 77.78%   Conservative: 2
Best Local Similarity: 84.09%   Mismatches: 0
Query Match: Indels: 0

DB: 13   Gaps: 0
US-10-617-443B-2_COPY_1_9 (1-9) x US-10-087-192-1735 (1-21562)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
    :::||||||||||:|||||||::|||||
Db   11652 ATCCACCTTGCCCTGCCCATCCAGTGCA 11678

RESULT 13
US-10-322-696-55
; Sequence 55, Application US/10322696
; Publication No. US20040166490A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Malandro, Marc
; TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
; FILE REFERENCE: 529452001200
; CURRENT APPLICATION NUMBER: US/10/322,696
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 247544
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(247544)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-696-55

Alignment Scores:
Pred. No.:   1.4e+05
Score: 37.00   Length: 247544
Percent Similarity: 88.89%   Matches: 7
Best Local Similarity: 77.78%   Conservative: 1
Query Match: 84.09%   Mismatches: 1
DB: 19   Indels: 0
      Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-322-696-55 (1-247544)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
    :::||||||||||||||:||||||||||
Db   23468 CTTCATCTTGCTTTTCCAGCTTCTGCA 23494

RESULT 14
US-10-189-099A-5
; Sequence 5, Application US/10189099A
; Publication No. US20040005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-10-189-099A-5

Alignment Scores:
Pred. No.:   26.5
Score: 36.00   Length: 34
Percent Similarity: 100.00%   Matches: 7
Best Local Similarity: 100.00%   Conservative: 7
      Mismatches: 0

```
Query Match:          81.82%       Indels:    0
DB:                   17           Gaps:      0

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-189-099A-5 (1-34)

Qy         1 MetHisLeuAlaLeuProAla 7
             |||||||||||||||||||||
Db        14 ATGCACCTGGCCCTTCCCGCC 34

RESULT 15
US-10-617-443B-5
; Sequence 5, Application US/10617443B
; Publication No. US20050019777A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT:    Qi, Jian-shen
; APPLICANT:    Chen, Cailin
; APPLICANT:    Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; CURRENT FILING DATE:  2003-07-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 5
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
;
US-10-617-443B-5

Alignment Scores:
Pred. No.:                         Length:         34
Score:              26.5           Matches:         7
Percent Similarity: 36.00          Conservative:    0
Best Local Similarity: 100.00%     Mismatches:      0
Query Match:        100.00%        Indels:          0
DB:                 81.82%         Gaps:            0
                    21

US-10-617-443B-2_COPY_1_9 (1-9) x US-10-617-443B-5 (1-34)

Qy         1 MetHisLeuAlaLeuProAla 7
             |||||||||||||||||||||
Db        14 ATGCACCTGGCCCTTCCCGCC 34

Search completed: June 30, 2005, 01:51:25
Job time : 52.1808 secs
```

```
                                    GenCore version 5.1.6
                                    Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:              June 29, 2005, 21:11:10 ; Search time 8.029215 Seconds
                                               (without alignments)
                                               1834.127 Million cell updates/sec

Title:               US-10-617-443B-2_COPY_1_9
Perfect Score:       44
Sequence:            1 MHIALPASA 9

Scoring table:       BLOSUM62
                     Xgapop 10.0 ; Xgapext    0.5
                     Ygapop 10.0 ; Ygapext    0.5
                     Fgapop  6.0 ; Fgapext    7.0
                     Delop   6.0 ; Delext     7.0

Searched:            1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:       2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:     Minimum Match 0%
                     Maximum Match 100%
                     Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_l/USPTO_spool/US10617443/runat_29062005_135932_26096/app_query.fasta_1.718
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rnl -MINMATCH=0.1 -LCQPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10617443 @CGN 1 1 93 @runat_29062005_135932_26096 -WAIT -DSFBLOCK=100 -LONGLOG
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -THREADS=1 -XGAPOP=10 -NCPU=6 -ICPU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :           Issued Patents NA:*
            1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	DB ID	Description
1	38	86.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
2	37	84.1	601	4	US-09-949-016-40948	Sequence 40948, A
3	37	84.1	601	4	US-09-949-016-40949	Sequence 40949, A
4	37	84.1	601	4	US-09-949-016-40950	Sequence 40950, A
5	37	84.1	601	4	US-09-949-016-40951	Sequence 40951, A
6	37	84.1	601	4	US-09-949-016-40952	Sequence 40952, A
7	37	84.1	601	1	US-09-949-016-107301	Sequence 107301,
8	37	84.1	2196	3	US-08-472-217-3	Sequence 3, Appli
9	37	84.1	2196	3	US-08-760-534A-3	Sequence 3, Appli
10	37	84.1	2196	4	US-09-336-757-3	Sequence 3, Appli
11	37	84.1	35688	4	US-09-949-016-16873	Sequence 16873, A
12	37	84.1	150780	4	US-09-949-016-14711	Sequence 14711, A
13	37	84.1	462589	4	US-09-949-016-12900	Sequence 12900, A
14	37	84.1	476044	4	US-09-949-016-12412	Sequence 12412, A
15	36	81.8	1854	4	US-09-620-312D-992	Sequence 992, App
16	36	81.8	5692	4	US-09-902-540-783	Sequence 783, App
17	36	81.8	9143	2	US-08-639-857-32	Sequence 32, Appl
18	36	81.8	9143	3	US-08-469-260A-390	Sequence 390, App
19	36	81.8	9143	3	US-08-469-260A-393	Sequence 393, App
20	36	81.8	9143	4	US-08-498-446-390	Sequence 390, App
21	36	81.8	9143	4	US-08-498-446-393	Sequence 393, App
22	36	81.8	9143	4	US-08-467-344A-390	Sequence 390, App
23	36	81.8	9143	4	US-08-467-344A-393	Sequence 393, App
24	36	81.8	9143	4	US-08-424-550B-390	Sequence 390, App
25	36	81.8	9143	4	US-08-424-550B-393	Sequence 393, App
26	36	81.8	18200	4	US-09-949-016-15661	Sequence 15661, A
27	36	81.8	18200	4	US-09-949-016-15660	Sequence 15660, A
28	36	81.8	19719	4	US-09-949-016-15662	Sequence 15662, A
29	36	81.8	19719	4	US-09-949-016-15663	Sequence 15663, A
30	36	81.8	150032	4	US-09-949-016-14321	Sequence 14321, A
31	35	79.5	601	4	US-09-949-016-76709	Sequence 76709, A
32	35	79.5	601	4	US-09-949-016-76710	Sequence 76710, A
33	35	79.5	601	4	US-09-949-016-145816	Sequence 145816,
34	35	79.5	601	4	US-09-949-016-146084	Sequence 146084,
35	35	79.5	601	4	US-09-949-016-146352	Sequence 146352,
36	35	79.5	601	4	US-09-949-016-160820	Sequence 160820,
37	35	79.5	601	4	US-09-949-016-160821	Sequence 160821,
38	35	79.5	601	3	US-08-340-533A-3	Sequence 162623,
39	35	79.5	1192	1	US-08-461-592B-3	Sequence 3, Appli
40	35	79.5	1192	2	US-09-270-767-12819	Sequence 12819, A
41	35	79.5	1362	4	US-08-965-879-12	Sequence 12, Appl
42	35	79.5	1469	4	US-09-865-879-12	Sequence 12, Appl
43	35	79.5	2188	4	US-09-620-312D-409	Sequence 409, App
44	35	79.5	2275	4	US-09-620-312D-410	Sequence 410, App
45	35	79.5	3224	3	US-08-965-729A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-103-840A-1/c
; Sequence 1, Application US/0910384OA
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.:   1.16e+06
Score: 38.00
Percent Similarity: 100.00%
Best Local Similarity: 77.78%
Query Match: 86.36%
DB: 3

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-103-840A-1 (1-4411529)

```
Qy          1 MetHisLeuAlaLeuProAlaSerAla 9
              :::||||||||:::||||||||||||||
Db    1987705 GTTCACTTAGCGATGCCGGCCGTCAGCT 1987679
```

RESULT 2
US-09-949-016-40948
; Sequence 40948, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40948
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40948

Qy   1 MetHisLeuAlaLeuProAlaSer 8
        |||||||||||||||||||||||||::||
Db   339 ATGCATCTAGCTTTGCCAAGTTCT 362

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-40948 (1-601)

Alignment Scores:
; Pred. No.:
Score: 164   Length: 601
             37.00   Matches: 7
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09%   Indel: 0
DB: 4   Gaps: 0

RESULT 3
US-09-949-016-40949
; Sequence 40949, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40949
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40949

Alignment Scores:
; Pred. No.:
Score: 164   Length: 601
             37.00   Matches: 7
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09%   Indel: 0
                 4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-40949 (1-601)

Qy   1 MetHisLeuAlaLeuProAlaSer 8
        |||||||||||||||||||||||::||
Db   340 ATGCATCTAGCTTTGCCAAGTTCT 363

RESULT 4
US-09-949-016-40950
; Sequence 40950, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40950
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40950

Alignment Scores:
; Pred. No.:
Score: 164   Length: 601
             37.00   Matches: 7
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 84.09%   Indel: 0
DB: 4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-40950 (1-601)

Qy   1 MetHisLeuAlaLeuProAlaSer 8
        |||||||||||||||||||||||::||
Db   344 ATGCATCTAGCTTTGCCAAGTTCT 367

RESULT 5
US-09-949-016-40951
; Sequence 40951, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40951
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-40951

Alignment Scores:
; Pred. No.:
Score: 164   Length: 601
             37.00   Matches: 7

Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 87.50%   Mismatches: 0
Query Match: 84.09%   Indels: 0
DB: 4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-40951 (1-601)

Qy   1 MetHisIleuAlaLeuProAlaSer 8
           |||||||||||||||||||||||::|||
Db   455 ATGCAtCtAGCTTTGCCAAGTTCT 478

RESULT 6
US-09-949-016-40952
; Sequence 40952, Application US/0994901 6
Db   Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40952
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human

US-09-949-016-40952

Alignment Scores:
Score: 164   Length: 601
Pred. No.:   Matches: 7
Percent Similarity: 37.00   Conservative: 1
Best Local Similarity: 100.00%   Mismatches: 1
Query Match: 87.50%   Indels: 0
DB: 84.09%   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-40952 (1-601)

Qy   1 MetHisIleuAlaLeuProAlaSer 8
           |||||||||||||||||||||||::|||
Db   479 ATGCAtCtAGCTTTGCCAAGTTCT 502

RESULT 7
US-09-949-016-107301
; Sequence 107301, Application US/0994901 6
Db   Patent No. 681233 9
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107301
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human

US-09-949-016-107301

Alignment Scores:
Pred. No.:
Score: 164   Length: 601
Percent Similarity: 37.00   Matches: 7
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 87.50%   Mismatches: 1
DB: 84.09%   Indels: 0
                                         Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-107301 (1-601)

Qy   1 MetHisIleuAlaLeuProAlaSer 8
           ::|||||||||||||||||||||||||||
Db   314 TTGCATTTGGCACTTCCTGCTTCT 337

RESULT 8
US-08-472-217-3/c
; Sequence 3, Application US/08472217
Db   Patent No. 5726058
; GENERAL INFORMATION:
; APPLICANT: Alanen-Kurki, Leena
; APPLICANT: Auvinen, Petri
; APPLICANT: Jaakkola, Panu
; APPLICANT: Jalkanen, Markku
; APPLICANT: Lepp , Sirpa
; APPLICANT: Mali, Markku
; APPLICANT: Vihinen, Tapani
; APPLICANT: Wirri, Anni
; TITLE OF INVENTION: Syndecan Stimulation Of Cellular
; TITLE OF INVENTION: Differentiation
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,217
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,427
; FILING DATE: 01-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,186
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1102.0050003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-472-217-3

Alignment Scores:

```
Pred. No.:
Score:                  666
Percent Similarity:     37.00%
Best Local Similarity:  100.00%
Query Match:            87.50%
DB:
                        Length:        2196
                        Matches:       7
                        Conservative:  1
                        Mismatches:    1
                        Indels:        0
                        Gaps:          0

US-10-617-443B-2_COPY_1_9 (1-9) x US-08-472-217-3 (1-2196)

Qy     1 MetHisLeuAlaLeuProAlaSer 8
         |||||:::||||||||||||||||||
Db  1143 ATGCATGTGGCACTCCCTGCCTCA 1120
```

RESULT 9
US-08-760-534A-3/c
Sequence 3, Application US/08760534A
Patent No. 6017727

```
GENERAL INFORMATION:
   APPLICANT: JALKANEN, MARKKU
   APPLICANT: JAAKKOLA, PANU
   APPLICANT: VIHINEN, TAPANI
   TITLE OF INVENTION: SYNDECAN ENHANCER ELEMENT AND SYNDECAN
   TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
   NUMBER OF SEQUENCES: 14
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
   STREET: 1100 NEW YORK AVENUE, SUITE 600
   CITY: WASHINGTON
   STATE: DC
   COUNTRY: US
   ZIP: 20005-3934
   COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/760,534A
   FILING DATE: 02-DEC-1996
   CLASSIFICATION: 435
   PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 08/206,186
   FILING DATE: 07-MAR-1994
   PRIOR APPLICATION DATA:
   APPLICATION NUMBER: PCT/FI93/00514
   FILING DATE: 01-DEC-1993
   ATTORNEY/AGENT INFORMATION:
   NAME: CIMBALA, MICHELE A.
   REGISTRATION NUMBER: 33,851
   REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
   TELECOMMUNICATION INFORMATION:
   TELEPHONE: (202) 371-2600
   TELEFAX: (202) 371-2540
   INFORMATION FOR SEQ ID NO: 3:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 2196 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: double
   TOPOLOGY: linear
   MOLECULE TYPE: DNA (genomic)
```

US-10-617-443B-2_COPY_1_9 (1-9) x US-08-760-534A-3 (1-2196)
Alignment Scores:

```
Pred. No.:
Score:                  666
Percent Similarity:     37.00%
Best Local Similarity:  100.00%
Query Match:            87.50%
DB:
                        Length:        2196
                        Matches:       7
                        Conservative:  1
                        Mismatches:    0
                        Indels:        0
                        Gaps:          0

US-10-617-443B-2_COPY_1_9 (1-9) x US-08-760-534A-3 (1-2196)

Qy     1 MetHisLeuAlaLeuProAlaSer 8
         |||||:::||||||||||||||||||
Db  1143 ATGCATGTGGCACTCCCTGCCTCA 1120
```

RESULT 10
US-09-336-757-3/c
Sequence 3, Application US/09336757
Patent No. 6492344

```
GENERAL INFORMATION:
   APPLICANT: JALKANEN, MARKKU
   APPLICANT: JAAKKOLA, PANU
   APPLICANT: VIHINEN, TAPANI
   TITLE OF INVENTION: SYNDECAN ENHANCER ELEMENT AND SYNDECAN
   TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
   NUMBER OF SEQUENCES: 14
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
   STREET: 1100 NEW YORK AVENUE, SUITE 600
   CITY: WASHINGTON
   STATE: DC
   COUNTRY: US
   ZIP: 20005-3934
   COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/336,757
   FILING DATE:
   CLASSIFICATION:
   PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 08/760,534
   FILING DATE:
   PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 08/206,186
   FILING DATE:
   PRIOR APPLICATION DATA:
   APPLICATION NUMBER: PCT/FI93/00514
   FILING DATE: 01-DEC-1993
   ATTORNEY/AGENT INFORMATION:
   NAME: CIMBALA, MICHELE A.
   REGISTRATION NUMBER: 33,851
   REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
   TELECOMMUNICATION INFORMATION:
   TELEPHONE: (202) 371-2600
   TELEFAX: (202) 371-2540
   INFORMATION FOR SEQ ID NO: 3:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 2196 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: double
   TOPOLOGY: linear
   MOLECULE TYPE: DNA (genomic)
```

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-336-757-3 (1-2196)
Alignment Scores:

```
Pred. No.:
Score:                  666
Percent Similarity:     37.00%
Best Local Similarity:  100.00%
Query Match:            87.50%
DB:
                        Length:        2196
                        Matches:       7
                        Conservative:  1
                        Mismatches:    0
                        Indels:        0
                        Gaps:          0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-336-757-3 (1-2196)

Qy     1 MetHisLeuAlaLeuProAlaSer 8
         |||||:::||||||||||||||||||
Db  1143 ATGCATGTGGCACTCCCTGCCTCA 1120
```

RESULT 11
US-09-949-016-16873
Sequence 16873, Application US/09949016
Patent No. 6812339

```
GENERAL INFORMATION:
   APPLICANT: VENTER, J. Craig et al.
   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
;     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;     FILE REFERENCE: CL001307
;     CURRENT APPLICATION NUMBER: US/09/949,016
;     CURRENT FILING DATE: 2000-04-14
;     PRIOR APPLICATION NUMBER: 60/241,755
;     PRIOR APPLICATION NUMBER: 60/237,768
;     PRIOR APPLICATION NUMBER: 60/231,498
;     PRIOR FILING DATE: 2000-10-03
;     PRIOR FILING DATE: 2000-10-20
;     PRIOR FILING DATE: 2000-09-08
;     NUMBER OF SEQ ID NOS: 207012
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 16873
;     LENGTH: 35688
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-16873

Alignment Scores:
Pred. No.:
Score:                     1.36e+04      Length:         35688
Percent Similarity:        88.89%        Matches:        7
Best Local Similarity:     77.78%        Conservative:   1
Query Match:               84.09%        Mismatches:     1
                                         Indels:         0
                                         Gaps:           0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-16873 (1-35688)

Qy     1 MetHisLeuAlaLeuProAlaSerAla 9
           |||||||||||||||||:::||||||||
Db 15904 ATGCACCTCACCCTCCCTAGTAGCGCT 15930


RESULT 12
US-09-949-016-14711
;     Sequence 14711, Application US/0994916
;     Patent No. 6812339
;     GENERAL INFORMATION:
;     APPLICANT: VENTER, J. Craig et al.
;     TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;     FILE REFERENCE: CL001307
;     CURRENT APPLICATION NUMBER: US/09/949,016
;     CURRENT FILING DATE: 2000-04-14
;     PRIOR APPLICATION NUMBER: 60/241,755
;     PRIOR APPLICATION NUMBER: 60/237,768
;     PRIOR APPLICATION NUMBER: 60/231,498
;     PRIOR FILING DATE: 2000-10-03
;     PRIOR FILING DATE: 2000-10-20
;     PRIOR FILING DATE: 2000-09-08
;     NUMBER OF SEQ ID NOS: 207012
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 14711
;     LENGTH: 150780
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-14711

Alignment Scores:
Pred. No.:
Score:                     6.4e+04       Length:         150780
Percent Similarity:        100.00%       Matches:        7
Best Local Similarity:     87.50%        Conservative:   1
Query Match:               84.09%        Mismatches:     0
                                         Indels:         0
                                         Gaps:           0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-14711 (1-150780)

Qy     1 MetHisLeuAlaLeuProAlaSer 8
           ||||||||||||||||||||||||
Db 42768 TTGCATTTGGCACTTCCTGCTTCT 42791


RESULT 13
US-09-949-016-12900/c
```

```
Qy     1 MetHisLeuAlaLeuProAlaSer 8

;     Sequence 12900, Application US/0994916
;     Patent No. 6812339
;     GENERAL INFORMATION:
;     APPLICANT: VENTER, J. Craig et al.
;     TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;     FILE REFERENCE: CL001307
;     CURRENT APPLICATION NUMBER: US/09/949,016
;     CURRENT FILING DATE: 2000-04-14
;     PRIOR APPLICATION NUMBER: 60/241,755
;     PRIOR APPLICATION NUMBER: 60/237,768
;     PRIOR APPLICATION NUMBER: 60/231,498
;     PRIOR FILING DATE: 2000-10-03
;     PRIOR FILING DATE: 2000-10-20
;     PRIOR FILING DATE: 2000-09-08
;     NUMBER OF SEQ ID NOS: 207012
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 12900
;     LENGTH: 462589
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-12900

Alignment Scores:
Pred. No.:
Score:                     2.09e+05      Length:         462589
Percent Similarity:        100.00%       Matches:        7
Best Local Similarity:     87.50%        Conservative:   1
Query Match:               84.09%        Mismatches:     0
                                         Indels:         0
                                         Gaps:           0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-12900 (1-462589)

Qy      1 MetHisLeuAlaLeuProAlaSer 8
            ||||||||||||||||||||||||
Db 431394 ATGCATCTAGCTTTGCCAAGTTCT 431371


RESULT 14
US-09-949-016-12412/c
;     Sequence 12412, Application US/0994916
;     Patent No. 6812339
;     GENERAL INFORMATION:
;     APPLICANT: VENTER, J. Craig et al.
;     TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;     TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;     FILE REFERENCE: CL001307
;     CURRENT APPLICATION NUMBER: US/09/949,016
;     CURRENT FILING DATE: 2000-04-14
;     PRIOR APPLICATION NUMBER: 60/241,755
;     PRIOR APPLICATION NUMBER: 60/237,768
;     PRIOR APPLICATION NUMBER: 60/231,498
;     PRIOR FILING DATE: 2000-10-03
;     PRIOR FILING DATE: 2000-10-20
;     PRIOR FILING DATE: 2000-09-08
;     NUMBER OF SEQ ID NOS: 207012
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 12412
;     LENGTH: 476044
;     TYPE: DNA
;     ORGANISM: Human
US-09-949-016-12412

Alignment Scores:
Pred. No.:
Score:                     2.15e+05      Length:         476044
Percent Similarity:        100.00%       Matches:        7
Best Local Similarity:     87.50%        Conservative:   1
Query Match:               84.09%        Mismatches:     0
                                         Indels:         0
                                         Gaps:           0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-949-016-12412 (1-476044)

Qy     1 MetHisLeuAlaLeuProAlaSer 8
```

```
Db    444850 ATGCATCTAGCTTTGCCAAGTTCT 444827
             |||||||||||||||||||::||||
```

RESULT 15

US-09-620-312D-992/c
Sequence 992, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 65696662e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 992
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (559)..(1710)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1854)
; OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-992

Alignment Scores:
Pred. No.:
Score: 874   Length: 1854
Percent Similarity: 36.00   Matches: 7
Best Local Similarity: 88.89%   Conservative: 1
Query Match: 77.78%   Mismatches: 1
DB: 81.82%   Indels: 0
                         4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x US-09-620-312D-992 (1-1854)

```
Qy     1 MetHisLeuAlaLeuProAlaSerAla 9
         |||||||||  |||||:|||||
Db   610 ATGCACCTGCAGCTGCCTGCAAACAGCT 584
```

Search completed: June 30, 2005, 01:41:07
Job time : 496.029 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 21:07:15 ; Search time 155.519 Seconds
                 (without alignments)
                 2202.808 Million cell updates/sec

Title: US-10-617-443B-2_COPY_1_9
Perfect score: 44
Sequence: 1 MHLALPASA 9

Scoring table: BLOSUM62
                 Xgapop   10.0 , Xgapext   0.5
                 Ygapop   10.0 , Ygapext   0.5
                 Fgapop   6.0 , Fgapext   7.0
                 Delop   6.0 , Delext   7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US10617443/runat_29062005_135932_26076/app_query.fasta_1.718
 -DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
 -USER=US10617443 @CGN 1_1_4385 @runat_29062005_135932_26076 -NCPU=6 -ICPU=3
 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                 1: gb_est1:*
                 2: gb_est2:*
                 3: gb_hc:*
                 4: gb_est3:*
                 5: gb_est4:*
                 6: gb_est5:*
                 7: gb_est6:*
                 8: gb_ss1:*
                 9: gb_ss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	41	93.2	150	4	BJ457822	BJ457822 BJ457822
2	41	93.2	181	4	BJ474645	BJ474645 BJ474645
3	41	93.2	202	4	BJ450407	BJ450407 BJ450407
c 4	41	93.2	234	6	CD662805	CD662805 UCRHV18_0
c 5	41	93.2	308	1	AV912195	AV912195 AV912195
c 6	41	93.2	314	6	CD662622	CD662622 UCRHV18_0
c 7	41	93.2	329	1	AV909564	AV909564 AV909564
c 8	41	93.2	332	1	AV911456	AV911456 AV911456
9	41	93.2	419	1	AV912337	AV912337 AV912337

c 10	41	93.2	427	1	AV910159	AV910159 AV910159
11	41	93.2	440	4	BJ461882	BJ461882 BJ461882
12	41	93.2	446	4	BJ455746	BJ455746 BJ455746
13	41	93.2	469	4	BJ458684	BJ458684 BJ458684
c 14	41	93.2	482	4	BJ448746	BJ448746 BJ448746
15	41	93.2	487	4	BJ447999	BJ447999 BJ447999
16	41	93.2	490	4	BJ451181	BJ451181 BJ451181
17	41	93.2	502	6	CD057466	CD057466 HO1SD11S
18	41	93.2	504	1	AJ434780	AJ434780 AJ434780
19	41	93.2	536	7	CK570032	CK570032 HO1SD11w
20	41	93.2	554	2	BE704839	BE704839 SC02.10h1
21	41	93.2	573	6	CB877974	CB877974 HP06O12T
22	41	93.2	590	2	BE492776	BE492776 WHE0564_F
23	41	93.2	600	2	BE705635	BE705635 Sc01.10b1
c 24	41	93.2	616	1	AV909381	AV909381 AV909381
c 25	41	93.2	619	6	BJ460127	BJ460127 BJ460127
c 26	41	93.2	628	6	CB878395	CB878395 HP08G11T
c 27	41	93.2	640	1	AV911273	AV911273 AV911273
28	41	93.2	739	2	BJ476744	BJ476744 BJ476744
c 29	41	93.2	648	6	CB877648	CB877648 HP05L04T
c 30	41	93.2	652	4	BJ452579	BJ452579 BJ452579
c 31	41	93.2	676	4	BJ453684	BJ453684 BJ453684
c 32	41	93.2	683	4	BJ458193	BJ458193 BJ458193
33	41	93.2	706	4	BJ461226	BJ461226 BJ461226
c 34	41	93.2	711	2	BE214392	BE214392 HV_CEb000
35	41	93.2	725	4	BF623796	BF623796 HVSMEa000
c 36	41	93.2	731	6	BG300006	BG300006 HVSMEa002
37	41	93.2	739	2	BF622309	BF622309 HVSMEa002
38	41	93.2	741	2	BF621719	BF621719 HVSMBa001
39	41	93.2	749	2	BF623605	BF623605 HVSMBa000
c 40	41	93.2	796	2	BF624602	BF624602 HVSMBa001
41	41	93.2	1101	9	CNS04VYJ	CNS04VYJ Terraodon
c 42	41	93.2	1175	2	BE421167	BE421167 HWM006.E0
c 43	39	88.6	439	4	AZ591722	AZ591722 1M0402A06
44	39	88.6	485	4	BI348039	BI348039 de5a01.x
45	39	88.6	510	1	AL638926	AL638926 AL638926

ALIGNMENTS

RESULT 1
BJ457822
LOCUS   BJ457822   150 bp   mRNA   linear   EST 23-MAY-2002
DEFINITION BJ457822 K. Sato unpublished cDNA library, cv. Akashinriki
            vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
            baak29g22 3', mRNA sequence.
ACCESSION   BJ457822
VERSION   BJ457822.1 GI:21136365
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 150)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..150
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak29g22"

ORIGIN

```
         /tissue_type="leaves"
         /dev_stage="vegetative stage"
         /clone_lib="K. Sato unpublished cDNA library, cv.
         Akashinriki vegetative stage leaves"
```

US-10-617-443B-2_COPY_1_9 (1-9) x BJ457822 (1-150)

Alignment Scores:

Pred. No.:   1
Score: 154   Length: 150
Percent Similarity: 100.00%   Matches: 8
Best Local Similarity: 88.89% Conservative: 1
Query Match: 93.18%   Mismatches: 0
DB: Indels: 0
                               Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
         :|||||||||||||||||||||||||||
Db   93 CTTCATTTGGCGCTGCCGGCGTCGGCT 119

RESULT 2
LOCUS   BJ474645   181 bp   mRNA   linear   EST 23-MAY-2002
DEFINITION   BJ474645 K. Sato unpublished cDNA library, cv. Haruna Nijo adult,
             heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA
             clone baa120g09 3', mRNA sequence.
ACCESSION   BJ474645
VERSION   BJ474645.1   GI:21153146
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 181)
AUTHORS   Sato,K., Saisho,D. and Takeda,K.
TITLE   Barley EST sequencing project in NIG and Okayama Univ
JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..181
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Haruna Nijo"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baa120g09"
                     /tissue_type="top three leaves"
                     /dev_stage="adult, heading stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo adult, heading stage top three leaves"
```

**US-10-617-443B-2_COPY_1_9 (1-9) x BJ474645 (1-181)**

Alignment Scores:

Pred. No.:            1
Score:             189         Length:            181
Percent Similarity: 100.00%    Matches:           8
Best Local Similarity: 88.89%  Conservative:      1
Query Match:        93.18%     Mismatches:        0
DB:                            Indels:            0
                               Gaps:              0

Qy    1 MetHisLeuAlaLeuProAlaSerAla 9
         :|||||||||||||||||||||||||||
Db   89 CTTCATTTGGCGCTGCCGGCGTCNGCT 115

---

**RESULT 3**
LOCUS        BJ450407/c          202 bp    mRNA    linear    EST 23-MAY-2002
DEFINITION   BJ450407 K. Sato unpublished cDNA library, cv. Akashinriki
             vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
             baak29g22 5', mRNA sequence.
ACCESSION    BJ450407
VERSION      BJ450407.1   GI:21129011
KEYWORDS     EST.
SOURCE       Hordeum vulgare subsp. vulgare
ORGANISM     Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE    1  (bases 1 to 202)
AUTHORS      Sato,K., Saisho,D. and Takeda,K.
TITLE        Barley EST sequencing project in NIG and Okayama Univ
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..202
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak29g22"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"
```

ORIGIN

US-10-617-443B-2_COPY_1_9 (1-9) x BJ450407 (1-202)

Alignment Scores:

Pred. No.:   1
Score: 213   Length: 202
Percent Similarity: 100.00%   Matches: 8
Best Local Similarity: 88.89% Conservative: 1
Query Match: 93.18%   Mismatches: 0
DB: Indels: 0
                               Gaps: 0

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
         :|||||||||||||||||||||||||||
Db   49 CTTCATTTGGCGCTGCCGGCGTCGGCT 23

RESULT 4
LOCUS   CD662805   234 bp   mRNA   linear   EST 23-JUN-2003
DEFINITION   CD662805 UCRHV18_03ag11_b1 Drought-stressed Dicktoo barley leaf
             epidermis cDNA library Hordeum vulgare subsp. vulgare cDNA clone
             UCRHV18_03ag11, mRNA sequence.
ACCESSION   CD662805
VERSION   CD662805.1   GI:32153419
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 234)
AUTHORS   Wahid,A., Close,T.J., Fenton,R.D., Wanamaker,S., Collura,K.,
             Feuerbacher,O., Kim,H.R., Kudrna,D., Wing,R. and Yu,Y.
TITLE   Drought-stressed barley leaf epidermis cDNA sequences
JOURNAL   Unpublished (2003)
COMMENT   Contact: Timothy J. Close
```

Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T7.

FEATURES
source    Location/Qualifiers
          1..234
          /organism="Hordeum vulgare subsp. vulgare"
          /mol_type="mRNA"
          /cultivar="Dicktoo"
          /sub_species="vulgare"
          /db_xref="taxon:112509"
          /clone="UCRHV18_03sg11"
          /tissue_type="lower leaf epidermis"
          /dev_stage="1-2 week seedlings"
          /lab_host="E. coli TJC121"
          /clone_lib="Drought-stressed Dicktoo barley epidermis cDNA library"
          /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
          Site_1: EcoRI; Site_2: XhoI; Seeds of barley (Hordeum
          vulgare L. cv. Dicktoo) were germinated in dishpans
          containing UCR-mix soil. Seedlings were kept in a growth
          chamber at 20C (day/night) and allowed to grow at 60-70%
          soil moisture content. After 3-4 days, the water was
          withheld in order to apply drought until the soil moisture
          content was reduced to 10-12%, which took another 3-4
          days. At the time of extraction of epidermis, leaf water
          and osmotic potentials and extracted lower epidermis
          osmotic potentials had dropped to 15.9%, 20.2% and 24.7%
          of controls, respectively, as determined using a vapor
          pressure osmometer (Model 5100C, Wescor, Inc., Logan, UT).
          Epidermal strips were quickly peeled off of seedlings and
          immediately submerged in liquid nitrogen. About 15-20 g of
          epidermal tissue was collected and used to extract total
          RNA. Total RNA was extracted using Concert Plant RNA
          Reagent (Invitrogen 12322-012). Poly(A)+ RNA was purified
          using PolyATtrack mRNA Isolation System IV (Promega). A
          cDNA library was made using a Uni-ZAP cDNA synthesis kit
          (Stratagene). A total of 1 million primary lambda cDNA
          clones were mass-excised in vivo to give a population of
          pBluescript SK(-) phagemids. All steps to this point were
          performed in the TJ Close lab at the University of
          California, Riverside by A. Wahid with some assistance
          from R.D. Fenton. Phagemids were plated on the TJC121 host
          strain, plasmid DNA purified, cDNA clones archived, and
          DNA sequences determined using the T7 primer (mainly 3'
          end reads) using an ABI3730 at the Arizona Genomics
          Institute, University of Arizona (Collura, Feuerbacher,
          Kim, Kudrna, Wing, Yu). Chromatogram files were
          transmitted to UC Riverside (by Yu), then processed at UC
          Riverside (by Wanamaker) using the HarvEST pipeline
          (http://harvest.ucr.edu) to remove vector and cloning
          oligo sequences and various contaminants, and to trim to a
          high quality region. Sequences that retained a phred 17
          region of at least 100 bases were deposited to GenBank."

ORIGIN

Alignment Scores:
Pred. No.:              251                Length:         234
Score:                  41.00              Matches:        8
Percent Similarity:     100.00%            Conservative:   1
Best Local Similarity:  88.89%             Mismatches:     0
Query Match:            93.18%             Indels:         0
                                           Gaps:           0

US-10-617-443B-2_COPY_1_9 (1-9) x CD662805 (1-234)

```
Qy    1 MetHisLeuAlaLeuProAlaSerAla 9
          :::|||||||||||||||||||||
Db  195 CTTCATTTGGCGCTCGCCGGCGTCGGCT 221
```

RESULT 5

LOCUS   AV912195   308 bp   mRNA   linear   EST 18-JAN-2002
DEFINITION AV912195 K. Sato unpublished cDNA library, cv. Akashinriki
            vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
            baakif21 3', mRNA sequence.
ACCESSION   AV912195
VERSION   AV912195.1 GI:18207972
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 308)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..308
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baakif21"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"
ORIGIN

Alignment Scores:
Pred. No.:   339   Length: 308
Score: 41.00   Matches: 8
Percent Similarity: 100.00%   Conservative: 1
Best Local Similarity: 88.89%   Mismatches: 0
Query Match: 93.18%   Indels: 0
                                           Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x AV912195 (1-308)

```
Qy    1 MetHisLeuAlaLeuProAlaSerAla 9
          :::|||||||||||||||||||||
Db  133 CTTCATTTGGCGCTCGCCGGCGTCGGCT 159
```

RESULT 6

LOCUS   CD662622   314 bp   mRNA   linear   EST 23-JUN-2003
DEFINITION UCRHV18_02cc08_b1 Drought-stressed Dicktoo barley epidermis cDNA
            library Hordeum vulgare subsp. vulgare cDNA clone UCRHV18_02cc08,
            mRNA sequence.
ACCESSION   CD662622
VERSION   CD662622.1 GI:32153043
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 314)
  AUTHORS   Wahid,A., Close,T.J., Fenton,R.D., Wanamaker,S., Collura,K.,
            Feuerbacher,O., Kim,H.R., Kudrna,D., Wing,R. and Yu,Y.
  TITLE   Drought-stressed barley leaf epidermis cDNA sequences
  JOURNAL   Unpublished (2003)
  COMMENT   Contact: Timothy J. Close

Column 1

```
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T7;
```

FEATURES
```
source          1..314
                /organism="Hordeum vulgare subsp. vulgare"
                /mol_type="mRNA"
                /cultivar="Dicktoo"
                /sub_species="vulgare"
                /db_xref="taxon:112509"
                /clone="UCRIV18_02cc08"
                /tissue_type="lower leaf epidermis"
                /dev_stage="1-2 week seedlings"
                /lab_host="E. coli TJC121"
                /clone_lib="Drought-stressed Dicktoo barley epidermis cDNA
                library"
                /note="Vector: lambda Uni-ZAP XR, excised phagemid;
                Site 1: EcoRI; Site 2: XhoI; Seeds of barley (Hordeum
                vulgare L. cv. Dicktoo) were germinated in dishpans
                containing UCR-mix soil. Seedlings were kept in a growth
                chamber at 20C (day/night) and allowed to grow at 60-70%
                soil moisture content. After 3-4 days, the water was
                withheld in order to apply drought until the soil moisture
                content was reduced to 10-12%, which took another 3-4
                days. At the time of extraction of epidermis, leaf water
                and osmotic potentials and extracted lower epidermis
                osmotic potentials had dropped to 15.9%, 20.2% and 24.7%
                of controls, respectively, as determined using a vapor
                pressure osmometer (Model 5100C, Wescor, Inc., Logan, UT).
                Epidermal strips were quickly peeled off of seedlings and
                immediately submerged in liquid nitrogen. About 15-20 g of
                epidermal tissue was collected and used to extract total
                RNA. Total RNA was extracted using Concert Plant RNA
                Reagent (Invitrogen 12322-012). Poly(A) RNA was purified
                using PolyATrack mRNA Isolation System IV (Promega). A
                cDNA library was made using a Uni-ZAP cDNA synthesis kit
                (Stratagene). A total of 1 million primary lambda cDNA
                clones were mass-excised in vivo to give a population of
                pBluescript SK(-) phagemids. All steps to this point were
                performed in the TJ Close lab at the University of
                California, Riverside by A. Wahid with some assistance
                from R.D. Fenton. Phagemids were plated on the TJC121 host
                strain, plasmid DNA purified, cDNA clones archived, and
                DNA sequences determined using the T7 primer (mainly 3'
                end reads) using an ABI3730 at the Arizona Genomics
                Institute, University of Arizona (Collura, Feuerbacher,
                Kim, Kudrna, Wing, Yu). Chromatogram files were
                transmitted to UC Riverside (by Yu), then processed at UC
                Riverside (by Wanamaker) using the HarvEST pipeline
                (http://harvest.ucr.edu) to remove vector and cloning
                oligo sequences and various contaminants, and to trim to a
                high quality region. Sequences that retained a phred 17
                region of at least 100 bases were deposited to GenBank."
```

ORIGIN

Alignment Scores:
Pred. No.:	1
Score:	346
Percent Similarity:	100.00%
Best Local Similarity:	88.89%
Query Match:	6

Length:	314
Matches:	8
Conservative:	1
Mismatches:	0
Indels:	0
Gaps:	0

US-10-617-443B-2_COPY_1_9 (1-9) x CD662622 (1-314)

```
      1 MetHisLeuAlaLeuProAlaSerAla 9
        |||||||||||||||||||||||||||
    177 CTTCATTTGGCGCTGCCCGCGTCGGCT 203
```

Column 2

RESULT 7
```
LOCUS           AV909564           329 bp    mRNA    linear    EST  18-JAN-2002
DEFINITION      AV909564 K. Sato unpublished cDNA library, cv. Akashinriki
                vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
                baak13e01 5', mRNA sequence.
ACCESSION       AV909564
VERSION         AV909564.1     GI:18205306
KEYWORDS        EST.
SOURCE          Hordeum vulgare subsp. vulgare
  ORGANISM      Hordeum vulgare subsp. vulgare
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Pooideae; Triticeae; Hordeum.
REFERENCE       1  (bases 1 to 329)
  AUTHORS       Sato,K., Saisho,D. and Takeda,K.
  TITLE         Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL       Unpublished (2002)
  COMMENT       Contact: Tadasu Shin-i
                Center For Genetic Resource Information
                National Institute of Genetics
                1111 Yata, Mishima, Shizuoka 411-8540, Japan
                Tel: 81-559-81-6856
                Fax: 81-559-81-6855
                Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..329
                /organism="Hordeum vulgare subsp. vulgare"
                /mol_type="mRNA"
                /cultivar="Akashinriki"
                /sub_species="vulgare"
                /db_xref="taxon:112509"
                /clone="baak13e01"
                /tissue_type="leaves"
                /dev_stage="vegetative stage"
                /clone_lib="K. Sato unpublished cDNA library, cv.
                Akashinriki vegetative stage leaves"
```

ORIGIN

Alignment Scores:
Pred. No.:	1
Score:	365
Percent Similarity:	100.00%
Best Local Similarity:	88.89%
Query Match:	1

Length:	329
Matches:	8
Conservative:	1
Mismatches:	0
Indels:	0
Gaps:	0

US-10-617-443B-2_COPY_1_9 (1-9) x AV909564 (1-329)

```
      1 MetHisLeuAlaLeuProAlaSerAla 9
        |||||||||||||||||||||||||||
    185 CTTCATTTGGCGCTGCCCGCGTCGGCT 159
```

Column 3

RESULT 8
```
LOCUS           AV911456           332 bp    mRNA    linear    EST  18-JAN-2002
DEFINITION      AV911456 K. Sato unpublished cDNA library, cv. Akashinriki
                vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
                baak13e01 3', mRNA sequence.
ACCESSION       AV911456
VERSION         AV911456.1     GI:18207233
KEYWORDS        EST.
SOURCE          Hordeum vulgare subsp. vulgare
  ORGANISM      Hordeum vulgare subsp. vulgare
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Pooideae; Triticeae; Hordeum.
REFERENCE       1  (bases 1 to 332)
  AUTHORS       Sato,K., Saisho,D. and Takeda,K.
  TITLE         Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL       Unpublished (2002)
  COMMENT       Contact: Tadasu Shin-i
                Center For Genetic Resource Information
```

National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

```
FEATURES
    source          1..332
                    /organism="Hordeum vulgare subsp. vulgare"
                    /mol_type="mRNA"
                    /cultivar="Akashinriki"
                    /sub_species="vulgare"
                    /db_xref="taxon:112509"
                    /clone="baak13e01"
                    /tissue_type="leaves"
                    /dev_stage="vegetative stage"
                    /clone_lib="K. Sato unpublished cDNA library, cv.
                    Akashinriki vegetative stage leaves"
ORIGIN
```

Alignment Scores:

Pred. No.:	368	Length:	332
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:		Gaps:	0

US-10-617-443B-2_COPY_1_9 (1-9) x AV911456 (1-332)

```
Qy    1 MetHisLeuValAlaLeuProAlaSerAla 9
          :::|||||||||||||||||||||||||
Db  131 CTTCATTTGGCGCTCCGGCGGTCGGCT 157
```

RESULT 9
AV912337

```
LOCUS       AV912337      419 bp    mRNA    linear    EST 18-JAN-2002
DEFINITION  AV912337 K. Sato unpublished cDNA library, cv. Akashinriki
            vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
            baak3g08 3', mRNA sequence.
ACCESSION   AV912337
VERSION     AV912337.1  GI:18208114
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM    Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 419)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6855
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES
    source          1..419
                    /organism="Hordeum vulgare subsp. vulgare"
                    /mol_type="mRNA"
                    /cultivar="Akashinriki"
                    /sub_species="vulgare"
                    /db_xref="taxon:112509"
                    /clone="baak3g08"
                    /tissue_type="leaves"
                    /dev_stage="vegetative stage"
                    /clone_lib="K. Sato unpublished cDNA library, cv.
                    Akashinriki vegetative stage leaves"
ORIGIN
```

Alignment Scores:

Pred. No.:	476	Length:	419
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:		Gaps:	0

US-10-617-443B-2_COPY_1_9 (1-9) x AV912337 (1-419)

```
Qy    1 MetHisLeuValAlaLeuProAlaSerAla 9
          :::|||||||||||||||||||||||||
Db  132 CTTCATTTGGCGCTGCCGGCGGTCGGCT 158
```

RESULT 10
AV910159/c

```
LOCUS       AV910159      427 bp    mRNA    linear    EST 18-JAN-2002
DEFINITION  AV910159 K. Sato unpublished cDNA library, cv. Akashinriki
            vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
            baak3g08 5', mRNA sequence.
ACCESSION   AV910159
VERSION     AV910159.1  GI:18205901
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM    Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 427)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6855
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES
    source          1..427
                    /organism="Hordeum vulgare subsp. vulgare"
                    /mol_type="mRNA"
                    /cultivar="Akashinriki"
                    /sub_species="vulgare"
                    /db_xref="taxon:112509"
                    /clone="baak3g08"
                    /tissue_type="leaves"
                    /dev_stage="vegetative stage"
                    /clone_lib="K. Sato unpublished cDNA library, cv.
                    Akashinriki vegetative stage leaves"
ORIGIN
```

Alignment Scores:

Pred. No.:	486	Length:	427
Score:	41.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	93.18%	Indels:	0
DB:		Gaps:	0

US-10-617-443B-2_COPY_1_9 (1-9) x AV910159 (1-427)

```
Qy    1 MetHisLeuValAlaLeuProAlaSerAla 9
          :::|||||||||||||||||||||||||
Db  289 CTTCATTTGGCGCTGCCGGCGGTCGGCT 263
```

RESULT 11
BJ461882

```
LOCUS       BJ461882      440 bp    mRNA    linear    EST 23-MAY-2002
DEFINITION  BJ461882 K. Sato unpublished cDNA library, cv. Akashinriki
            vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
            baak21b12 3', mRNA sequence.
ACCESSION   BJ461882
```

VERSION   BJ461882.1 GI:21140391
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 440)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..440
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak21b12"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 502   Length: 440
Percent Similarity: 41.00   Matches: 8
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 88.89%   Mismatches: 0
DB: 93.18%   Indels: 0
                                4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ461882 (1-440)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
           :::||||||||||||||||||||||||||
Db   102 CTTCATTTGGCGCTGCCCGGCGTCGGCT 128

RESULT 12
LOCUS   BJ455746
DEFINITION   BJ455746 K. Sato unpublished cDNA library, cv. Akashinriki
             vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
             baak18a21 3', mRNA sequence.
ACCESSION   BJ455746
VERSION   BJ455746.1 GI:21134312
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 446)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..446

                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak18a21"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 509   Length: 446
Percent Similarity: 41.00   Matches: 8
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 88.89%   Mismatches: 0
DB: 93.18%   Indels: 0
                                4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ455746 (1-446)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
           :::||||||||||||||||||||||||||
Db   104 CTTCATTTGGCGCTGCCCGGCGTCGGCT 130

RESULT 13
LOCUS   BJ458684
DEFINITION   BJ458684 K. Sato unpublished cDNA library, cv. Akashinriki
             vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
             baak33a14 3', mRNA sequence.
ACCESSION   BJ458684
VERSION   BJ458684.1 GI:21137220
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 469)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
  COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..469
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak33a14"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 538   Length: 469
Percent Similarity: 41.00   Matches: 8
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 88.89%   Mismatches: 0
DB: 93.18%   Indels: 0
                                4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ458684 (1-469)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
        :::|||||||||||||||||||||||||||
Db 126 CTTCATTTGGCGCTGGCGCCGGCCGTCGGCT 152

RESULT 14
BJ448746/c
LOCUS   BJ448746   482 bp   mRNA   linear   EST 23-MAY-2002
DEFINITION   BJ448746 K. Sato unpublished cDNA library, cv. Akashinriki
             vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
             baak2lb12 5', mRNA sequence.
ACCESSION   BJ448746
VERSION   BJ448746.1 GI:21127368
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 482)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..482
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak2lb12"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"

ORIGIN

Alignment Scores:
Pred. No.:
Score: 555   Length: 482
Percent Similarity: 41.00   Matches: 8
Best Local Similarity: 100.00% Conservative: 1
Query Match: 88.89%   Mismatches: 0
DB: 93.18%   Indels: 0
                    4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ448746 (1-482)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
        :::|||||||||||||||||||||||||||
Db 338 CTTCATTTGGCGCTGGCGCCGGCCGTCGGCT 312

RESULT 15
BJ447999/c
LOCUS   BJ447999   487 bp   mRNA   linear   EST 23-MAY-2002
DEFINITION   BJ447999 K. Sato unpublished cDNA library, cv. Akashinriki
             vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone
             baak18a21 5', mRNA sequence.
ACCESSION   BJ447999
VERSION   BJ447999.1 GI:21126629
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. vulgare
  ORGANISM   Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.

REFERENCE   1 (bases 1 to 487)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source   1..487
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Akashinriki"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="baak18a21"
                     /tissue_type="leaves"
                     /dev_stage="vegetative stage"
                     /clone_lib="K. Sato unpublished cDNA library, cv.
                     Akashinriki vegetative stage leaves"

ORIGIN

Alignment Scores:
Pred. No.:
Score: 561   Length: 487
Percent Similarity: 41.00   Matches: 8
Best Local Similarity: 100.00% Conservative: 1
Query Match: 88.89%   Mismatches: 0
DB: 93.18%   Indels: 0
                    4   Gaps: 0

US-10-617-443B-2_COPY_1_9 (1-9) x BJ447999 (1-487)

Qy   1 MetHisLeuAlaLeuProAlaSerAla 9
        :::|||||||||||||||||||||||||||
Db 343 CTTCATTTGGCGCTGGCGCCGGCCGTCGGCT 317

Search completed: June 30, 2005, 01:27:45
Job time : 161.519 secs

THIS PAGE BLANK (USPTO)

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 19:59:10 ; Search time 7513.54 Seconds
(without alignments).
2153.984 Million cell updates/sec

Title: US-10-617-443B-2
Perfect score: 1679
Sequence: 1 MHLALPASAGLHQLSSPRYK.........LEVRRGNDDLLFSIAPBVVM 334

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 2422607955 residues

Total number of hits satisfying chosen parameters: 9416466.

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10617443/runat_29062005_135931_26061/app_query.fasta_1.718.
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=-1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10617443 @CGN 1 1 3890 @runat 29062005 135931 26061 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1635	97.4	1797	6	AX338509	AX338509 Sequence
2	1635	97.4	2040	6	BD260123	BD260123 Matrix-re
3	1635	97.4	2543	6	AY280665	AY280665 Homo sapi
4	1635	97.4	2554	9	AY040094	AY040094 Homo sapi

ALIGNMENTS

No.	Score	%	Length	DB	ID	Description
5	1635	97.4	2585	9	BC034390	BC034390 Homo sapi
6	1635	97.4	2586	9	BC035717	BC035717 Homo sapi
7	1536	91.5	2446	10	AY156509	AY156509 Mus muscu
8	1524	90.8	1338	9	CQ879976	CQ879976 Sequence
9	1463.5	87.2	2525	10	AY037300	AY037300 Mus muscu
10	1450.5	86.4	2007	5	CR353995	CR353995 Gallus ga
11	1450.5	86.4	2176	5	BX929862	BX929862 Gallus ga
12	1238.5	73.8	2080	5	BC078402	BC078402 Danio rer
13	1117	66.5	1953	9	AY280666	AY280666 Homo sapi
14	1102.5	65.7	819	9	CQ716317	CQ716317 Sequence
15	1057	63.0	1897	10	AY280664	AY280664 Mus muscu
16	1042.5	62.1	1113	4	AF097707	AF097707 Bos tauru
17	1034.5	61.6	1894	9	AX747270	AX747270 Sequence
18	1034.5	61.6	1894	9	AK091944	AK091944 Homo sapi
19	1034.5	61.6	1900	9	AK092476	AK092476 Homo sapi
20	1034.5	61.6	2029	9	HSSERIPRT	HSSERIPRT
21	1034.5	61.6	2036	9	CQ728266	CQ728266 Sequence
22	1034.5	61.6	2036	9	D87258	D87258 Homo sapien
23	1034.5	61.6	2075	9	E12965	E12965 cDNA encodi
24	1029.5	61.3	1169	9	AF097709	AF097709 Homo sapi
25	1029.5	61.3	2043	9	BC031082	BC031082 Homo sapi
26	1027.5	61.2	2012	10	AF193370	AF193370 Rattus no
27	1027.5	61.2	2036	6	AR095627	AR095627 Sequence
28	1027.5	61.2	2065	10	BC081767	BC081767 Rattus no
29	1026.5	61.1	1777	5	BX934040	BX934040 Gallus ga
30	1026.5	61.1	2005	10	AF173369	AF173369 Mus muscu
31	1026.5	61.1	2007	10	AK128916	AK128916 Mus muscu
32	1023.5	61.0	2023	10	BC013516	BC013516 Mus muscu
33	1006.5	59.9	1705	6	BD194550	BD194550 Human nuc
34	1006.5	59.9	1705	6	AX015404	AX015404 Sequence
35	1000.5	59.6	2010	10	AF172994	AF172994 Mus muscu
36	989.5	58.9	2055	9	BC074069	BC074069 Danio rer
37	981.	58.5	1534	6	AX202090	AX202090 Sequence
38	925	55.1	1362	6	BC022646	BC022646 Mus muscu
39	893.5	53.2	915	6	BX950476	BX950476 Gallus ga
40	890	53.0	1544	6	CQ783299	CQ783299 Sequence
41	890	53.0	1544	6	BD127576	BD127576 Sequence
42	890	53.0	1544	6	AK075205	AK075205 Homo sapi
43	890	53.0	1553	6	BD063243	BD063243 Secreted
44	888	52.9	1521	6	BC057765	BC057765 Mus muscu
45	883	52.6	1625	10	AF164513	AF164513 Mus muscu

RESULT 1
AX338509

LOCUS   AX338509   1797 bp   DNA   linear   PAT 09-JAN-2002
DEFINITION Sequence 28 from Patent WO0183775.
ACCESSION   AX338509
VERSION   AX338509.1 GI:18128931
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Delegeane,A.M., Lal,P., Hafalia,A., Patterson,C., Walia,N.K.,
            Kearney,L., Tribouley,C.M., Khan,F.A., Yao,M.G., Baughn,M.R.,
            Azimzai,Y., Elliott,V.S., Nguyen,D.B., Gandhi,A.R., Yang,J.,
            Hernandez,R., Policky,J.L., Lu,D.A., Reddy,R. and Tang,Y.T.
TITLE   Proteases
JOURNAL   Patent: WO 0183775-A 28 08-NOV-2001;
            Incyte Genomics, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1797
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7474343CB1"
ORIGIN
Alignment Scores:

```
Pred. No.:              1.066-130
Score:                  1635.00
Percent Similarity:     100.00%        Length:          1797
Best Local Similarity:  100.00%        Matches:          325
Query Match:            97.38%         Conservative:       0
                                       Mismatches:         0
                                       Indels:             0
                                       Gaps:               0

US-10-617-443B-2 (1-334) x AX338509 (1-1797)

Qy   10  GlyIleuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu   29
         GGTGTCCACCAGCTGAGCAGCCCCGGCTACAAGTTCAATTGCTTCATTGCTGTGGTGGAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  565  GGTGTCCACCAGCTGAGCAGCCCCGGCTACAAGTTCAATTGCTTCATTGCTGTGGTGGAG  624

Qy   30  LysIleAlaProAlaValValHisIleLeuGluLeuPheLeuArgHisProLeuPheGlyArg   49
         AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGTGTTTGGGCGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  625  AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGTGTTTGGGCGC  684

Qy   50  AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleLeuThr   69
         AACGTGCCCCTGTCCAGCGGTTCAGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  685  AACGTGCCCCTGTCCAGCGGTTCAGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC  744

Qy   70  AsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnGlnLeuLysValGln   89
         AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCGCCGCGCCAGGTCGCCAGGTGCAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  745  AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCGCCGCGCCAGGTCGCCAGGTGCAG  804

Qy   90  LeuGlnAsnGlyAspSerTyrGluIleAlaThrIleLysTyrSerSerAspIle  109
         CTACAGAATGGGGACTCCTATGAGGCCGCCACCATCAAAGACATCGGACATT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  805  CTACAGAATGGGGACTCCTATGAGGCCGCCACCATCAAAGACATCGGACATT  864

Qy  110  AlaThrIleLysIleHisProLysGlyLysLeuProValLeuLeuLeuGlyHisSerAla  129
         GCCACCATCAAGATCCACCCCAAGAAAAAGTCCCCTGTGTTGTTGCTGGGTCACTCGGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  865  GCCACCATCAAGATCCACCCCAAGAAAAAGTCCCCTGTGTTGTTGCTGGGTCACTCGGC  924

Qy  130  AspLeuArgProGlyGluPheValAlaAlaIleGlySerProPheAlaLeuGlnAsnThr  149
         GACCTGCGGCCTGGGGAGTTTGTGGCGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  925  GACCTGCGGCCTGGGGAGTTTGTGGCGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA  984

Qy  150  ValThrThrGlyIleValSerThrThrAlaGlnAsnArgGluGlyLeuLeuGlyArgAsp  169
         GTGACAACGGGCATCGTCAGCACTGCCCAGAACCGGGAGGGCTGGGCCCTCCGGGAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  985  GTGACAACGGGCATCGTCAGCACTGCCCAGAACCGGGAGGGCTGGGCCCTCCGGGAC  1044

Qy  170  SerAspMetAsnTyrYrIleGlnThrAspAlaIleLeuIleLeuAsnTyrGlyAsnGlyPro  189
         TCCGACATGGACTACTACATCCAGACGGATGCCATCAACTACGGGAACTCCGGGGACCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1045  TCCGACATGGACTACTACATCCAGACGGATGCCATCAACTACGGGAACTCCGGGGACCA  1104

Qy  190  LeuValAsnLeuAspGlyValLeuIleLeuAsnThrLeuLysValThrAlaGlyIle  209
         CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1105  CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC  1164

Qy  210  SerPheAlaIleProSerSerAspArgIleThrArgPheLeuThrGluProGlnLysGln  229
         TCCTTTGCCATCCCCTCAGACCGGATCACACGGTTCCTCACAGAGTTCCAAAGACAAGCAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1165  TCCTTTGCCATCCCCTCAGACCGGATCACACGGTTCCTCACAGAGTTCCAAAGACAAGCAG  1224

Qy  230  IleLysAspTrpLysTyrGlyArgPheIleGlyIleLeuArgMetArgThrIleThrProSerLeu  249
         ATCAAAGACTGGAAGAAGGCCTTCATCGGCATACGGATGAGGACTATACACACAAGCCTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1225  ATCAAAGACTGGAAGAAGGCCTTCATCGGCATACGGATGAGGACTATACACACAAGCCTG  1284

Qy  250  ValAlaAspGluLeuLysValAlaSerArgAsnProLeuAspProGluValValSerGlyIleLeuTyrVal  269
         GTGGATGAGCTGAAGGTCGCTAGCCGCAACCCGCTGGACCCGGAAATTATGTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1285  GTGGATGAGCTGAAGGTCGCTAGCCGCAACCCGCTGGACCCGGAAATTATGTG  1344

Qy  270  GlnGluValAlaProSerProSerGlnAsnGlyIleGlnIleGlnAsnGlyAspIleIle  289
         CAAGAGGTTGCGCCGAGTCCATCTCAGAATGGGATTCAGATTCAGAACGGTGACATCATC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1345  CAAGAGGTTGCGCCGAGTCCATCTCAGAATGGGATTCAGATTCAGAACGGTGACATCATC  1404

Qy  290  ValLysValAsnGlyArgProLeuValAspSerGluGlnGlnLeuGluGlnValAlaValLeuThr  309
         GTCAAGGTCAACGGGCGTCCTCTAGTGGACTCGGAGCAGCAGCTGGAGCAGGTCGCTGTCC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1405  GTCAAGGTCAACGGGCGTCCTCTAGTGGACTCGGAGCAGCAGCTGGAGCAGGTCGCTGTCC  1464

Qy  310  GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspSerLeuPheSerIleAla  329
         GAGTCTCCTCTTCTACTGGAGGTCCGCCGGGGCAACGACGACTCCCTCTTCAGCATCGCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1465  GAGTCTCCTCTTCTACTGGAGGTCCGCCGGGGCAACGACGACTCCCTCTTCAGCATCGCA  1524
```

```
RESULT 2
BD260123
BD260123

Qy  330  ProGluValValMet  334
         CCTGAGGTGGTCATG
         |||||||||||||||
Db 1525  CCTGAGGTGGTCATG  1539
```

LOCUS   BD260123   2040 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Matrix-remodeling genes.
ACCESSION   BD260123
VERSION   BD260123.1 GI:33069893
KEYWORDS   JP 2002527054-A/14.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2040)
  AUTHORS   Walker,M.G., Volkmuth,W. and Klingler,T.M.
  TITLE   Matrix-remodeling genes
  JOURNAL   Patent: JP 2002527054-A 14 27-AUG-2002;
            INCYTE PHARMACEUTICALS INC
COMMENT   OS   Homo sapiens (human)
            PN   JP 2002527054-A/14
            PD   27-AUG-2002
            PF   06-OCT-1999 JP 2000575891
            PR   06-OCT-1999 JP 09/162289
            PI   MICHAEL G WALKER, WAYNE VOLKMUTH, TOD M KLINGLER PC
            C12N15/09, A61K38/00, A61K38/28, A61K48/00, A61P1/04, PC
            A61P3/10,
            A61P9/10, A61P17/02, A61P19/02, A61P35/00, C07K14/47, C07K16/18, PC
            C12N1/15,
            C12N1/19, C12N1/21, C12N5/10, C12Q1/68, G01N33/15, G01N33/50, G01N33/ PC
            53,
            G01N33/53//A61K31/711, A61K35/76, C12N15/00, C12N5/00, A61K37/02, PC
            A61K37/26,
            A61K37/54
            CC   Incyte ID NO.: 2814981CB1
            FH   Key   Location/Qualifiers
            FT   source   1..2040
            FT   /organism='Homo sapiens (human)'.

FEATURES   Location/Qualifiers
     source   1..2040
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

ORIGIN

```
Alignment Scores:
Pred. No.:              1.25e-130
Score:                  1635.00
Percent Similarity:     100.00%        Length:          2040
Best Local Similarity:  100.00%        Matches:          325
Query Match:            97.38%         Conservative:       0
                                       Mismatches:         0
                                       Indels:             0
                                       Gaps:               0

US-10-617-443B-2 (1-334) x BD260123 (1-2040)

Qy   10  GlyIleuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu   29
         GGTGTCCACCAGCTGAGCAGCCCCGGCTACAAGTTCAATTGCTTCATTGCTGTGGTGGAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72  GGTGTCCACCAGCTGAGCAGCCCCGGCTACAAGTTCAATTGCTTCATTGCTGTGGTGGAG  131

Qy   30  LysIleAlaProAlaValValHisIleLeuGluLeuPheLeuArgHisProLeuPheGlyArg   49
         AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGTGTTTGGGCGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132  AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGTGTTTGGGCGC  191

Qy   50  AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleLeuThr   69
         AACGTGCCCCTGTCCAGCGGTTCAGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  192  AACGTGCCCCTGTCCAGCGGTTCAGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC  251

Qy   70  AsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnGlnLeuLysValGln   89
         AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCGCCGCGCCAGGTCGCCAGGTGCAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  252  AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCGCCGCGCCAGGTCGCCAGGTGCAG  311
```

RESULT 3

```
LOCUS       AY280665               2543 bp    mRNA    linear   PRI 23-MAY-2003
DEFINITION  Homo sapiens pregnancy-related serine protease HTRA3 mRNA, complete
            cds; alternatively spliced.
ACCESSION   AY280665
VERSION     AY280665.1  GI:31044217
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2543)
  AUTHORS   Nie,G.Y., Hampton,A., Li,Y., Findlay,J.K. and Salamonsen,L.A.
  TITLE     Identification and cloning of two isoforms of human
            high-temperature requirement factor A3 (HtrA3), characterization of
            its genomic structure and comparison of its tissue distribution
            with HtrA1 and HtrA2
  JOURNAL   Biochem. J. 371 (Pt 1), 39-48 (2003)
   MEDLINE  22533278
   PUBMED   12513693
REFERENCE   2  (bases 1 to 2543)
  AUTHORS   Nie,G., Hampton,A., Li,Y., Findlay,J.K. and Salamonsen,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-2003) Prince Henry's Institute of Medical
            Research, 246 Clayton Rd., Melbourne, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1..2543
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4p16.1"
                     /tissue_type="heart"
     CDS             193..1554
                     /codon_start=1
                     /note="alternatively spliced"
                     /product="pregnancy-related serine protease HTRA3"
                     /protein_id="AAP42282.1"
                     /db_xref="GI:31044218"
                     /translation="MQARALLLAALLALLALAREPPAAPCPARCDVSRCPSPRCPGGYV
                     PDLCNCCLVCAAASEGEPCCGPLDSPCGESLECVRGLCRCRWSHAVCGTDGHTYANVCA
                     LQAASRRALQLSGTPVRQLQKGACGPLGLHLQLSSPRYKFNFIADVHVHIELF
                     LRHPLFGRNVPLSGSSGFIMSEAGLIITNAHVVSSNSAAPGRQQLKVQLONGDSYEAT
                     IKDIDKKSDIATIKIHPKKKLPVLLLGHSADLRPGEFVVAIGSPPALQNTVTTGIVST
                     AQRGRELGLRDSDMDYIQTDAIINYNGNSGGPLVNLDGEVIGINTLKVTAGISAIPS
                     DRIRFLTEFQDKQIKDWKKRFIGIRMRITPSLVDELKASNPDFPEEVSSGIYQEVA
                     PNSPSQRGGIQDGDIIVKVNGRPLVDSSELQEAVLITESPLLEVRRGNDDLLFSTAPE
                     VVM"
ORIGIN
```

US-10-617-443B-2 (1-334) x AY280665 (1-2543)

Alignment Scores:

```
Pred. No.:                              Length:          2543
Score:            1635.00               Matches:          325
Percent Similarity: 100.00%             Conservative:       0
Best Local Similarity: 100.00%          Mismatches:         0
Query Match:        97.38%              Indels:             0
                    9                   Gaps:               0
```

```
Qy         10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 GGTCTCCACCAGCTGAGCAGCAGCCCCGCCTACAAAGTTCAACTTCATTGCTGACGTGGTGGAG 636

Qy         30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AAGATCGCACCAGCCGTGGTCCACATAGAGCTGTTCCTGAGACACCCGCTGTTTGGCCGC 696

Qy         50 AsnValProLeuSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 AACGTGCCCCTGTCCAGGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 756

Qy         70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCCGGGCAGGCAGCTCAAGGTGCAG 816

Qy         90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAspIle 109
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCGGACATT 876

Qy        110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGTTGGGGTCACTCGGCC 936

Qy        130 AspLeuArgProGlyGluPheValValAlaIleGlySerProProAlaLeuGlnAsnThr 149
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        937 GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCCCAGCACTGCAGAACACA 996

Qy        150 ValThrThrGlyIleValSerThrAlaGlnArgGlyArgGluLeuGlyLeuArgAspAsp 169
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        997 GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGGCAGGGAGCTGGGCCTCCGGGAC 1056

Qy        170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
```

```
RESULT 4
LOCUS       AY040094    2554 bp    mRNA    linear    PRI 05-JUN-2002
DEFINITION  Homo sapiens serine protease HTRA3 mRNA, complete cds.
ACCESSION   AY040094
VERSION     AY040094.2  GI:21326482
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2554)
AUTHORS     Southan,C. and Punia,P.K.
TITLE       Direct Submission
JOURNAL     Submitted (14-JUN-2001) Computational Biology, Gemini Genomics, 162
            Science Park, Milton Rd., Cambridge CB4 0GH, UK
REFERENCE   2  (bases 1 to 2554)
AUTHORS     Southan,C.
TITLE       Direct Submission
JOURNAL     Submitted (05-JUN-2002) Proteome Discovery, Oxford Glycosciences,
            The Forum, 86 Milton Park, Abingdon, Oxon OX14 4RY, UK
REMARK      Sequence update by submitter
COMMENT     On Jun 5, 2002 this sequence version replaced gi:14700055.
FEATURES             Location/Qualifiers
     source          1..2554
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="between D4S412 and D4S1601"
     CDS             222..1583
                     /note="S1C serine peptidase family; similar to bacterial
                     heat-shock protease HtrA"
                     /codon_start=1
                     /product="serine protease HTRA3"
                     /protein_id="AAK71475.2"
                     /db_xref="GI:21326483"
                     /translation="MQARALLLAALLAALALARPPAAPCPARCDVSRCPSPRCPGGYV
                     PDLCNCCLVCAASEGBPCGGPLDSPCGGESILECVRGLCRCRWSHAVCGTDGHTYANVCA
                     LQAASRRALQLSGTPVRQLOKQACPLGLHQLSSPRYKFNFIADVVEKIAPAVVHIELF
                     LRHPLFGRNVPLSSSGSGPIMSEAGLIITNAHVVSSNSAAPGRQQLKVQLONGDSYEAT
                     IKDIDKKSDIATIKIHPKKKLPVLLLGHSADLRGEFVVAIGSPFALQNTVTGIVST
                     AQREGRELGLRDSDMDYIQTDAIINRGNSGGPLVNLDGEVIGINTLKVFAGISPAIPS
                     DRITRFLTEFQDKQIRDWKRFIGIRMRTITPSLVDELKASNPDFPEVSSGIIYQEBVA
                     PNSPSQRGGIQDGDIIVKVNGRPLVDSSELQEAVLTESPLLLEVRRGNDDLLFSIAPE
                     VVM"
```

ORIGIN

Alignment Scores:
```
Pred. No.:             1.67e-130      Length:         2554
Score:                 1635.00        Matches:         325
Percent Similarity:    100.00%        Conservative:      0
Best Local Similarity: 100.00%        Mismatches:        0
Query Match:           97.38%         Indels:            0
DB:                    9              Gaps:              0
```

US-10-617-443B-2 (1-334) x AY040094 (1-2554)

```
Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  606 GGTCTCCACCAGCTGAGCAGCCCCGGCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 665

Qy   30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  666 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGTGTTTGGGCGC 725

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  726 AACGTGCCCCTGTCCAGCGGGTTCGGCCTTCATCAGGAGCCCGGCCTGATCATCACC 785

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  786 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGCCAGCAGCTCAAGGTGCAG 845

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAspIle 109
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  846 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCGGACATT 905

Qy  110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  906 GCCACCATCAAGATCCATCCCAAGAAAAAAGCTCCCTGTGTTGTTGCTGGGCACTCCGCC 965

Qy  130 AspLeuArgGlyGluPheValAlaAlaIleGlySerProPheAlaLeuGlnAsnThr 149
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  966 GACCTGCGGGCCTGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 1025

Qy  150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1026 GTGACAACCGGCATCGTCAGCACTGCCCAGAGGGAGGGCCGTGGGCCTTGGGCTGGGGAC 1085

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnArgGlyAsnSerGlyGlyPro 189
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1086 TCCGGACATGGACTACATCCAGACGGATGCCATCATCAACACTACGGGGAACTCCGGGGGACCA 1145

Qy  190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValPheAlaGlyIle 209
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1146 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCTTCGCCGGCATC 1205

Qy  210 SerProAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1206 TCCTTGCCATCCCCTTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 1265

Qy  230 IleLysAspTrpLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1266 ATCAAAGACTGGAAGAGAGCCTTCATCGGCATACGGATGCGGACGATCACCAAGCCCTG 1325

Qy  250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleIleTyrVal 269
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1326 GTGGATGAGCTGAAAGGCCAGCAGAACCCGGACTTCCCAGAGGTGTCAGCAGTGGAATTTATTTG 1385

Qy  270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGlyAspIleIle 289
```

ORIGIN
```
Db 1537 CCTGAGGTGGTGCATG 1551
Qy  330 ProGluValValMet 334
```

```
Db    1386 CAAGAGGTTGCGCCGAATTCACCTTCTCAGAGAGCGGGCCATCCAAGAATGGTGACATCATC 1445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnLeuThr 309

Db    1446 GTCAAGGTCAACGGGCGCCCTCCTTCAGTGGACTCGAGTCGAGCTGCAGGAGCCGGTCGACC 1505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspLeuLeuPheSerIleAla 329

Db    1506 GAGTCTCCTCCTACTGGAGGTGGCGCGCGGGGGGGACAGACCTCCTCTTCAGCATCGCA 1565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     330 ProGluValValMet 334

Db    1566 CCTGAGGTGGTCATG 1580
```

FEATURES   Location/Qualifiers

<table>
<tr><td>Db</td><td>1386</td><td>...</td></tr>
</table>

RESULT 5

BC034390

LOCUS   BC034390   2585 bp   mRNA   linear   PRI 29-JUN-2004

DEFINITION   Homo sapiens serine protease HTRA3, mRNA (cDNA clone MGC:35339
    IMAGE:5180585), complete cds.

ACCESSION   BC034390

VERSION   BC034390.1   GI:21706740

KEYWORDS   MGC.

SOURCE   Homo sapiens (human)

ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 2585)

AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
    Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
    Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
    Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
    Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
    Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
    Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
    Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
    Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
    McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
    Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
    Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
    Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
    Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
    Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
    Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
    Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
    Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE   Generation and initial analysis of more than 15,000 full-length
    human and mouse cDNA sequences

JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED   12477932

REFERENCE   2 (bases 1 to 2585)

AUTHORS   Strausberg,R.

TITLE   Direct Submission

JOURNAL   Submitted (02-JUL-2002) National Institutes of Health, Mammalian
    Gene Collection (MGC), Cancer Genomics Office, National Cancer
    Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
    USA

COMMENT   Contact: MGC help desk
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: Life Technologies, Inc.
    cDNA Library Preparation: Life Technologies, Inc.
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Baylor College of Medicine Human Genome
    Sequencing Center
    Center code: BCM-HGSC
    Web site: http://www.hgsc.bcm.tmc.edu/cdna/
    Contact: amg@bcm.tmc.edu
    Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
    Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
    A.N., Gibbs, R.A.

REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
    through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
    Series: IRAK Plate: 50 Row: b Column: 22
    This clone was selected for full length sequencing because it
    passed the following selection criteria: Hexamer frequency ORF
    analysis.

    clone distribution: MGC clone distribution information can be found

FEATURES   Location/Qualifiers
    source   1..2585
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="MGC:35339 IMAGE:5180585"
       /tissue_type="Brain, Lung, Testis, adult, pooled whole"
       /clone_lib="NIH_MGC_115"
       /lab_host="DH10B"
       /note="Vector: pCMV-SPORT6"

    gene   1..2585
       /gene="HTRA3"
       /db_xref="LocusID:94031"

    CDS   205..1566
       /gene="HTRA3"
       /codon_start=1
       /product="serine protease HTRA3"
       /protein_id="AAH34390.1"
       /db_xref="GI:21706741"
       /db_xref="LocusID:94031"
       /translation="MQARLLLAALLALALAREPPAAPCPARCDVSRCPSPGGYV
       PDLCNCLLVCAASEGEPCGPLLDSPCGESLECVRLRCRCWSHAVCGTDGHTYANVCA
       LQAASRRALLQLSGTPVRQLQKGGLPGIQHLGSSPRYKFNFIADVVEKIAPAVVHIELF
       LRHPLFGRNVPLSSGGSGFIMSEAGLIITNAHVVSSNSAAPGRQQLKVQLQNGDSYEAT
       IKDIDKKSDIATIKIHPKKKLPVLLLGHSADLRPGEFVVAIGSPFALQNTVTTGIVST
       AQRGREIGLRDSDMDYIQTDAIINYGNSGGPLVNLDGSVIGINTLKVTAGISSPAIPS
       DRIRFLTEFQDKQIKDWKRPIGIRMRTITPSLVDELKASNPDFPEVSSGIYVQEVA
       PNSPSQRGGIQDGDIIVKVNGRPLVDSSELQEAVLTESPLLLEVRRGNDDLLFSIAPE
       VVM"

ORIGIN

US-10-617-443B-2 (1-334) x BC034390 (1-2585)

Alignment Scores:
Pred. No.:
Score: 1.69e-130
Percent Similarity: 1635.00
Best Local Similarity: 100.00%
Query Match: 97.38%

Length: 2585
Matches: 325
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0
DB:

```
Qy    10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   589 GGTCTCCACGTGGTCGTCAGCAACAGTGCGTGCCCACCAGGAGCCAGCTCAAAGGTGGAG 648

Qy    30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   649 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 708

Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   709 AACGTGCCCTGTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 768

Qy    70 AlaThrHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   769 AATGCCCACGTGGTGTCCAACAGTGCGTGCCCCGGAGGCAGCAGCTCAAGGTGCAG 828

Qy    90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAspIle 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   829 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCGGACATT 888

Qy   110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   889 GCCACCATCAAGATCCATCCAAAGAAAAAGCTCCCTGTGTTGTTGTTGGGTCACTCGGCC 948

Qy   130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   949 GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 1008
```

```
Qy    150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1009 GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGCCGAGGGGAGCTGGGCCTCCGGGAC 1068

Qy    170 SerAspMetAspTyrIleGlnThrAspAlaIleLeuAsnTyrGlyAsnSerGlyGlyPro 189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1069 TCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGACCA 1128

Qy    190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1129 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACCCTCAAGGTCACGGCTGGCATC 1188

Qy    210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1189 TCCTTTGCCATCCCCTCAGACCGCATCACCCGGTTCCTCACAGAGTTCCAAGACAAGCAG 1248

Qy    230 IleLysAspTrpLysGlyArgPheIleGlyIleHisMetArgThrIleThrProSerLeu 249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1249 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATACGGCATCAGGAGAGGATCACACCAAGCCTG 1308

Qy    250 ValAspGluLeuLysAlaSerAsnProAsnProPheGluValValSerSerGlyIleTyrVal 269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1309 GTGGATGAGCTGAAGGCCAGCAACCCGAACCCGTTCGAGGTCGTGTCCAGTGGAATTTATGTG 1368

Qy    270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnIleAsnGlyAspIleIle 289
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1369 CAAGAGGTTGCGCCCAATTCACCTTCACAGAGAGGGGGCCATCCAAGATAGTGGACATCATC 1428

Qy    290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1429 GTCAAGGTCAACGGGCGTCCTCTAGTGGACTCGAGTGAGCTGCAGGAGGCCGTGCTGACC 1488

Qy    310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1489 GAGTCTCCTCTACTGGAGGTCCGGCGGGGCAACGACGACCTTCTCTTCAGCATCGCA 1548

Qy    330 ProGluValValMet 334
          |||||||||||
Db   1549 CCTGAGGTGGTCATG 1563
```

RESULT 6
LOCUS   BC035717
DEFINITION Homo sapiens serine protease HTRA3, mRNA (cDNA clone MGC:46243
            IMAGE:5588608), complete cds.
ACCESSION   BC035717
VERSION   BC035717.1 GI:23273037
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2586)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length

            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 2586)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 79 Row: e Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 24475740.
FEATURES   Location/Qualifiers
     source   1..2586
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:46243 IMAGE:5588608"
                     /tissue_type="Ovary, pooled from 3 adults"
                     /clone_lib="NIH_MGC_125"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene   1..2586
                     /gene="HTRA3"
                     /db_xref="LocusID:94031"
                     /note="HTRA3"
     CDS   205..1566
                     /gene="HTRA3"
                     /codon_start=1
                     /product="serine protease HTRA3"
                     /protein_id="AAH35717.1"
                     /db_xref="GI:23273038"
                     /db_xref="LocusID:94031"
                     /translation="MQRARLLAALAALAALAREPPAAPCPARCDVSRCPSPRCPGGYV
                     PDLCNCCLVCAASEGEPCGGPLDSPCGESLECVRGICRCRWSHAVCGTDGHTYANVCA
                     LQAASRALQLSGTPVRQLQGACPLGLHLSSPRYKFNFIADVEKIAPAVVHIELF
                     LRHPLFGRNVPLSSGSGFIMSEAGLIITNAHVVSSNSAAPGRQQLKVQLQNGDSYEAT
                     IKDIDKKSDIATKIHPKKLPVLLGHSADLRPGEFVVAIGSPFALQNTVTTGIVST
                     AQREGRELGLRDSDMDYIQTDAIINYGNSGGPLVNLDGEVIGINTLKVTAGISFAIPS
                     DRITRFLTEFQDKQIKDWKKRFIGIRMRTITPSLVDELKASNPDPFEVSSGIYVQEVA
                     PNSPSQRGGIQGDIIVKVNGRPLVDSSELQEAVLTESPLLLEVRRGNDDLLFSIAPE
                     VVM"
ORIGIN

Alignment Scores:
Pred. No.:   1.69e-130   Length: 2586
Score: 1635.00   Matches: 325
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 97.38%   Indels: 0
DB: Gaps: 0

```
LOCUS       AY156509                2446 bp    mRNA    linear   ROD 01-OCT-2003
DEFINITION  Mus musculus pregnancy-related serine protease mRNA, complete cds.
ACCESSION   AY156509
VERSION     AY156509.1  GI:37359367
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2446)
  AUTHORS   Nie,G., Li,Y., Findlay,J.K. and Salamonsen,L.A.
  TITLE     Identification and cloning of a novel pregnancy-related serine
            protease (PRSP)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2446)
  AUTHORS   Nie,G., Li,Y., Findlay,J.K. and Salamonsen,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2002) Prince Henry's Institute of Medical
            Research, 246 Clayton Road, Melbourne, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1..2446
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Swiss"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="placenta; uterus"
     CDS             127..1506
                     /codon_start=1
                     /note="PRSP"
                     /product="pregnancy-related serine protease"
                     /protein_id="AAO17289.1"
                     /db_xref="GI:37359368"
                     /translation="MQARALLPATLAILAILATLAVALAREPPAARCDVSRCPSPR
                     CPGGYVPDLCNCCLVCAASEGEPGCRPLDSPCGDSLECVRGVCRCRWTHVCGIDGHT
                     YADVCALQAASRRALQVSGTPVRQLQKGACPSGLHQLTSPRYKENFIADVVEKIAPAV
                     VHIELFLRHPLFGRNVPLSSSGSGPIMSRAGLIVTNAHVVSSSSTASGRQQLKVQLQNG
                     DAVEATIQDIDKKSDIATIVIHPKKKLPVLLLGHSADLRPGEFVVAIGSPFALQNTVT
                     TGIVSTAQRDGKELGLRDSDMDYIQTDAIINYGNSGGPLVNLDGSEVIGINTLKVAAGI
                     SFAIPSDRITRFLSEFQNKHVKDWKKKRIIGIRMRTITPSLVEELKAANDFPPAVSSGI
                     YVQEVVPNSPSQRGGIQDGDIIVKVNGRPLADSSELQEAVLNESSLLLEVRRGNDDLL
                     FSIIPEVVM"
ORIGIN
```

```
US-10-617-443B-2 (1-334) x AY156509 (1-2446)

Alignment Scores:
Pred. No.:           4.94e-122
Score:               1536.00          Length:        2446
Percent Similarity:  96.63%           Matches:       302
Best Local Similarity: 92.64%         Conservative:  13
Query Match:         91.48%           Mismatches:    11
DB:                  10               Indels:        0
                                      Gaps:          0
```

```
US-10-617-443B-2 (1-334) x AY156509 (1-2446)

Qy     9 AlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValVal 28
          :::::||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db   526 TCTGGTCTCCACCAGCTGACCAGTCCGCGGTACAAGTTCAACTTCATCGCCGATGTGGTG 585

Qy    29 GlyLeuValAlaProAlaAlaValHisIleGluLeuPheLeuArgHisProLeuPheGly 48
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   586 GGAGAAGAATTCGGCCGCCAGCTGTGGTCCACATAGAGCTCTTTCTGAGACACCCCCTG 645

Qy    49 ArgAsnValProLeuSerSerSerGlySerGlyProIleMetSerArgGluAlaGlyLeuIle 68
          ||||||||||||:||||||||:||||||||||||||||||||||||||||||||||||
Db   646 CGGAATGTGCCGCTGTCCAGTGGTCGGGGTTCAACATGTCAGAAGCCGGTTTGATCGTC 705

Qy    69 ThrAsnAlaHisValValSerSerSerAsnSerAlaAlaAlaProGlyArgGlnGlnLeuLysVal 88
          ||||||||||||||||||:||||||||||||||||||||||||||||||||||:||
Db   706 ACCAACGCCCACGTGGTCTCCAGCTCCAGCAACTGCCTCCGGCCAGCAGCTGAAGGTG 765

Qy    89 GlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAsp 108
          ||||||||||||:||||||||:::|||||:::||||||||||||||||||||||||||
Db   766 CAGCTGCAGAATGGGGATGTCCTATGAGGCCACCATCAAGGACATCGACAAGAAGTCGGAC 825
```

```
RESULT 8
CQ879976
LOCUS       CQ879976                1338 bp    DNA     linear   PAT 11-OCT-2004
DEFINITION  Sequence 40 from Patent WO200408324.
ACCESSION   CQ879976
VERSION     CQ879976.1  GI:54038832
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Sakamoto,T. and Takeda,S.
  TITLE     Btc-interacting proteins and use thereof
  JOURNAL   Patent: WO 200408324-A 40 30-SEP-2004;
            Takeda Chemical Industries, Ltd. (JP)
FEATURES             Location/Qualifiers
     source          1..1338
                     /organism="Mus musculus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10090"
ORIGIN
.
Alignment Scores:
```

```
US-10-617-443B-2 (1-334) x CQ879976 (1-1338)
DB:

Pred. No.:              2.43e-121      Length:           1338
Score:                  1524.00        Matches:          300
Percent Similarity:     96.01%         Conservative:     13
Best Local Similarity:  92.02%         Mismatches:       13
Query Match:            90.77%         Indels:           0
                                       Gaps:             0
```

```
Qy    9 AlaGlyTyrLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAlaProValVal 28
        :::|||||:||||||::::|||||::::|||||||||||||||:::
Db  358 TCTGGCTCCCACCCAGCTGACCAGTCCCGGGTACAAAGTTCAACTTCATCGCCGCTCCCTCACTCA 417

Qy   29 GluLysIleAlaProAlaValValHisIleGlyLeuPheLeuArgHisProLeuPheGly 48
        |||||:|||||||:||||||||||||:|||:||||||||||||||||:::
Db  418 GAGAAGATTGCCCCAGCTGTGGTCCACATAGGCCTCTTTCTGAGACACCCCTTGTTCGG 477

Qy   49 ArgAsnValProLeuSerSerGlySerGlyGlyPheMetSerGlyValAlaGlyVal 68
        ||||||||||||||||||||||||||||::::|
Db  478 CGGAAATGTCCCTGTCTCCAGTGGCTCGGGTGGCTTCATGTCAGGAGCGGGTTGATCGTC 537

Qy   69 ThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnAspAsp 88
        ||||||||||||||||||||||:::|
Db  538 ACCAACGCCCACGTGGTCTCCAGCTCCGCCGCTCCAGGACACCGACAAGAAGTCGGAC 597

Qy   89 GlnLeuGlnAsnGlyAspSerThrTyrGlyGluValThrIleThrProSer 108
        ||||||||||||||||:::|
Db  598 CAGCTGCAGAATGGGGATAGCACCTATGGCGAGGTCACCATCACGCCAAGT 657

Qy  109 IleAlaThrIleLysIleHisProLysLysLeuProValLeuLeuLeuGlyHisSer 128
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  658 ATTGCCACGATTGTAATCCACCCCAAGAAAAAGCTCCCTGTGTTGCTGCTGGGTCACTCA 717

Qy  129 AlaAspLeuArgProGlyGluPheValValAlaLeuIleGlySerProPheAlaLeuGlnAsn 148
        ||||||||||||||||||||||||||||||||:::|||||||||||||||
Db  718 GCAGACCTGCGCCCTGGCGAGTTCGTGGTGGCCATCGGCAGCCCCTTTGCCCTGCAGAAC 777

Qy  149 ThrValThrThrGlyIleValSerThrArgLeuGlnAsnArgGluValGlyArgGlyLeuArg 168
        |||||||||||||:::|||||||||||||||||||||||||||||||||
Db  778 ACCGTGACAACGGGCATTGTCAGCACTCGCCTGCAAAATAGGGAAGTTGGTCGCGGACTC 837

Qy  169 AspSerAspMetAspTyrIleGlnThrAspAlaAlaIleLeuIleLeuAsnTyrGlyAsnSerGly 188
        ||||||||||||||||||:::|||||||||||||||||||||||
Db  838 GACTCAGACATGGACTATATCCAGACCGATGCCGCCATCCTCATCCTCAATTACGGGAACTCAGGAGGA 897

Qy  189 ProLeuValAsnLeuAspGlyGluValIleIleGluAsnThrLeuLysValThrArgAlaGly 208
        ||||||||||||||||||||||:::|||||||||||||
Db  898 CCCCTGGTGAACCTGGATGGGGAGGTCATCAGCGAGAACACGCTCAAGGTTGCAACTGGC 957

Qy  209 IleSerPheAlaLeuIleLeuProSerAspAsnGlyIleThrArgPheLeuThrGluProSerLys 228
        ||||||||||||:||||||:::|||||||||||||||||||||
Db  958 ATCTCCTTTGCCATCCCCTCAGATCGCAATGGCATCACGCGCTTCCTCACTGAGTTCCAAAAACAG 1017

Qy  229 GlnIleLysAspTrpLysTyrGlyArgPheIleGluArgMetArgThrIleThrProSer 248
        |||||||||||||||||||||||||:||||:|||||||||||||||||
Db 1018 CAGGTGAAAGACTGGAAGTACGGCCGCTTCATTGAGCGGATGCGACCATCACGCCAAGT 1077

Qy  249 LeuValIleAspGlyLeuLysSerAsnProAspPheProGluValSerGlyIleTyr 268
        ||||||||||||||:::|||||||||||||||||||||||||||||
Db 1078 TTGGTGATCGACGGCCTGAAGAGCAACCCAGACTTTCCAGAGGTTTCAGGGATCTAT 1137

Qy  269 ValGlnAsnGlyValAlaProAsnSerProSerGlnPheGlyLeuIleAspGlyIle 288
        |||:||||||||||||||||||||:::||||||||||||||||||||
Db 1138 GTTCAGAAGGGTGGTTCCCCAAATCACCCTTCTCAGAGGAGGCACATC 1197

Qy  289 IleValLysValAsnGlyArgProLeuValAspSerSerGlyIleGlyAlaValLeu 308
        ||||||:::||||||||||||||||||||||||||||||||||||
Db 1198 ATCGTCAAAGTCAATGGCCGCCCCCTGGGCGACTCCAGGGCAGTCCTG 1257

Qy  309 ThrGluSerProLeuLeuLeuGlyLeuValArgGlyAsnAspProLeuLeuPheSerIle 328
        ||||||
Db 1258 AACGAGTCTTCACTTCTTCTCTGGCTTGTCCGGGGAGGCCAATGATGAATCTCCTTCTTCAGCATC 1317
```

RESULT 9

```
Qy    329 AlaProGluValMet 334
          ||||||||||||||
Db   1318 ATCCCTGAGGTGGTCATG 1335
```

LOCUS   AY037300   2525 bp   mRNA   linear   ROD 07-JUL-2001
DEFINITION Mus musculus toll-associated serine protease (Tasp) mRNA, complete
            cds.
ACCESSION   AY037300
VERSION   AY037300.1 GI:14626470
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2525)
  AUTHORS   Matsuguchi,T. and Yoshikai,Y.
  TITLE   TASP, a novel mammalian serine protease
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2525)
  AUTHORS   Matsuguchi,T. and Yoshikai,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-MAY-2001) Medicine, Nagoya University, 65
            Tsurumai-cho, Nagoya, Aichi 466-8550, Japan
FEATURES   Location/Qualifiers
     source   1..2525
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
     gene   1..2525
                     /gene="Tasp"
     CDS   220..1602
                     /gene="Tasp"
                     /note="TASP"
                     /codon_start=1
                     /product="toll-associated serine protease"
                     /protein_id="AAK70226.1"
                     /db_xref="GI:14626471"
                     /translation="MQARALLPATLALTLAVTLAVALAREPPAAPCPARCDVSRCPSPR
                     CPGGYVPDLCNCCLVCAASEGEPCGRPLDSPCGDSLECVRGVCRCRWTHTVCGTDGHT
                     YADVCALQAAASRRALQVSGTPVRQLQKACPSGLHQLTSPRYKFNFIADVVEKSIADAS
                     VHIELFLRHPLFGRNVPLSSGSGFIMSEAGLIVTNAHVVSSSTASGRQQLKVQLQNG
                     DAYEATIQDIDKKSDIATIVIHPKKLPLVLLGHSADLRPGEFVVAIESPLCPAEHRD
                     KIVSTAORTGKELGLRDSDMDYIQTDAIINYGNSGGPLVNLDGEVIGINTLKVAAGI
                     SFAIPSDRITRFLSEFQNKHVKDWKKRPIGIRMRTITPSLVEELKAANPDFPAVSSGI
                     YVQEVVPNSPSQRGGIQDGDIIVKVNGRPLADSSELQEAVLNESSLLAAEVRRGNDDL
                     LFSIIPEVVM"
```

ORIGIN

```
US-10-617-443B-2 (1-334) x AY037300 (1-2525)

Alignment Scores:
Pred. No.:
Score:                 8.57e-116
Percent Similarity:      94.82%
Best Local Similarity:   90.85%
Query Match:             87.16%
DB:                      10

                        Length:        2525
                        Matches:        298
                        Conservative:    13
                        Mismatches:      15
                        Indels:           3
                        Gaps:             1
```

```
Qy      9 AlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValVal 28
          ::||||:||||||||||||||||||||||::::|||||:||||||||:||||||||||||
Db    799 ACCAAGCGCCCAGGTGGTCTCCAGCCACTGCCTTCCCGGCCCGCAGCCAGCTGAAGGTG 858

Qy     29 GluLysIleAlaProAlaValHisIleGluLeuPheLeuArgHisProLeuPheGly 48
          ||||||||:||||||||:|||||||||||||||:|||||||||||||||||||||
Db    679 GAGAAGAATTGCGCCACGCTGTGGTCCACATAGAGTCTTTTCTGAGACACCCCTGTTTGGC 738

Qy     49 ArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIle 68
          |||||||||||||||||:|||||||||||||:||||||||||||||||||||:||:
Db    739 CGGAATGTGCCGCTGTCCAGTGGCTCGGGCTTCATCATGTCAGAAGCCGGTTTGATCGTC 798
```

RESULT 10

```
Qy    327 IleAlaProGluValMet 334
          ||||||||||||||
Db   1578 CATCATCCCTGAGGTGGTCATG 1599
```

LOCUS   CR353995   2007 bp   mRNA   linear   VRT 11-MAR-2004
DEFINITION Gallus gallus finished cDNA, clone ChEST625h5.
ACCESSION   CR353995
VERSION   CR353995.1 GI:45425889
KEYWORDS   .
SOURCE   Gallus gallus (chicken)
  ORGANISM Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 2007)
  AUTHORS   Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
            Croning,M.D.R., Davies,R.M., Francis,M.D., Graham,D.V.,
            Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
            Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
```

```
Qy     69 ThrAsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnGlnLeuLysVal 88
          :|||||:|||||||||||||:|||||||||||||||||||||||||:||||||||||||||
Db    799 ACCAACGCCCAGGTGGTCTCCAGCCACTGCCTTCCCGGCCCGCAGCCAGCTGAAGGTG 858

Qy     89 GlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleGlnAspIleAspLysLysSerAsp 108
          ||||||||||||||||:||||||::||||||||||||||||||||:::||||||||||||
Db    859 CAGCTGCAGAATGGGAGATGCCTTATGAGGCCACCATCCAGGACATCGACAAGAAGTCGGAC 918

Qy    109 IleAlaThrIleLysIleHisProLysLysLeuProLeuValLeuLeuGlyHisSer 128
          ||||||||||||||:||||||||||||||||:|||||:||||||||||||||:|||
Db    919 ATTGCCACGATTGTAATCCACCCAAGAAAAAGCTCCCTGTGGTTGCTGCTGGGTCACTCA 978

Qy    129 AlaAspLeuArgProGlyGluPheValValAlaIleGluSer--ProPheAlaLeuGlnAsp 148
          |||||:||||||||||||||||||||||||:|||||||     |||||||||||||
Db    979 GCAGACCTGCGCCCTGGCAGAGTTCGTGGTGGCCATCGAGAGTCCCCTTTGCCCTGCAGAA 1038

Qy    149 nThrValThrThrGlyLeuValSerThrAlaGlnAlaGlyGluGlyLeuGlyHisSer 168
          ||||||:|::||||||||||:||||||||||||||||||:|||||||||||||||:|
Db   1039 CACCGTGACAACGTGT-ATTGTCAGCACTGCCCCACCGAAGAAGAGCTGGGTCTCCG 1097

Qy    169 gAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGly 188
          |||||||||:||||||||||||||||||||:|||||||||||:|||||||||||||||
Db   1098 GGACTGACAATGGACTATATCCAGACCGATGCCATCATCAATTACGGGAACTCAGGAGG 1157

Qy    189 yProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGl 208
          |||||||||||:||||||||||||:|:||||:|||||||:||||||||||:|||||||
Db   1158 ACCCCTGGTGAACCTGGATGGCGAGGTCATCGGCATCAACACGCTCAAGGTTGCAGGTGG 1217

Qy    209 yIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLy 228
          |||||:||||||||||||||||:|||||||||||||||||||||||||||:::|||||
Db   1218 CATCTCCTTTGCCATCCCATCAGATAGGATCACACGCTTCCTCACTGAGTTCCAAAACA 1277

Qy    229 sGlnIleLysAspTrpLysArgPheIleGlyIleLysMetArgThrIleThrProSe 248
          ::|||||||||||||||||||||||||||||:||:||||||:||||||||||||:
Db   1278 GCATGTGAAAGACTGGAAGAGACGCGCTTCATTGGCATCAAGATGCGGACCATCACGCCAAG 1337

Qy    249 rLeuValAspGluGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTy 268
          ||||||||||||::||||||||||||||||||||:||||||:||||||||||||||:|||||
Db   1338 TTTGGTGGACGAAGAACTGAAAGCCGCCAACCCAGACTTTCCAGAGGTGTCATCAGGAATATA 1397

Qy    269 rValGlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIl 288
          :|||||||||||||||||:|||||||||||||||||||||||||||||||||:||
Db   1398 TGTTCAAGAGGTGGTTCCCAATTCACCTTCTCAGAGAGGGCATCCAGGATGGGGACAT 1457

Qy    289 eIleValLysValAsnGlyArgProLeuValAlaAspSerSerGluLeuGlnLeuAlaValIle 308
          ||||||||||||||||||||:|||||||||:|||||||||||||||||||||||:||||||
Db   1458 CATCGTCAAAGTCAATGGCCGCCCCCTGGTCGCTGACAGCTCCGAGCTGCAGGAGGCAGTCCT 1517

Qy    309 uThrGluSerProLeuLeu---LeuGluValArgArgGlyAsnAspAspLeuLeuPhe 327
          :|||||||::|||||||||   |||||||||||||||||||||:||||||||||||
Db   1518 GAACGAGTCTTCACTCCTAGCTGCCGAAGTGCGGCGAGGCAATGAGGATCTCCTCTTCAG 1577
```

```
TITLE
JOURNAL   Tickle,C. and Wilson,S.A.
          Direct Submission
          Submitted (09-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: chickest@bms.umist.ac.uk

COMMENT   BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
          sequencing project.
          This sequence is from the
          BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
          from a library constructed by Elizabeth Bosch.
          from RNA extracted from small intestine, normalised, and poly
          A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector.
          Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host:
          Escherichia coli DH10B.

FEATURES
     source   Location/Qualifiers
              1..2007
              /organism="Gallus gallus"
              /mol_type="mRNA"
              /strain="Compton Line 15I"
              /db_xref="taxon:9031"
              /clone="ChEST625h5"
              /clone_lib="CSEOQCHN56"
              /dev_stage="adult"

ORIGIN

US-10-617-443B-2 (1-334) x CR359995 (1-2007)

Alignment Scores:
  Pred. No.:
  Score:                    8.3e-115      Length:       2007
                            1450.50       Matches:       291
  Percent Similarity:        91.15%       Conservative:   18
  Best Local Similarity:     85.84%       Mismatches:     21
  Query Match:               86.39%       Indels:          9
                                          Gaps:            1
```

```
Qy    5 LeuProAlaSerAla----------------------GlyLeuHisGlnLeuSer  15
        |||                                ||| ||| ||
Db  175 CTGCCCGCCGGCCTGGCTCAGGTGCTGCAGAAAGGAGCCATGGGCTCACCTACAGTCCAGC 234

Qy   16 SerProArgThrTyrLeuSerPheAsnPheIleAlaAspValValGluLysIleAlaProAlaVal 35
        || ||| ||
Db  235 AGTCCAAGAACAAATTCAACTTCATAGCAGATGTGGTGGAAAAGATTGCACCTGCAGTT 294

Qy   36 ValHisIleGluLeuProLeuArgHisProLeuPheGlyArgAsnValProLeuSerSer 55
        ||| |||
Db  295 GTGCACATTGAACTTCCTCAGGCACCCTCTGTTTGGTCGAAATGTCCCACTTTCCAGT 354

Qy   56 GlySerGlyPheIleMetSerGluAlaGlyLeuLeuIleIleThrAsnAlaHisValValSer 75
        ||| |||
Db  355 GGATCTGGGTTTATTATGTCTGAATTCGGTTTAATTGTGACCAAATGCACACGTGGTGTCA 414

Qy   76 SerAsnSerAlaAlaProGlyArgGlnGlnLeuGlnLeuAsnGlyAspSer 95
        ::|||
Db  415 AGTACCAATGCTATATCAGGGAGGACAACAGTTGAAAGTCCAGCTACAGAATGGAGATACA 474

Qy   96 TyrGluAlaThrIleLysAspIleAlaThrIleLysIleHis 115
        ::|||
Db  475 TATGAAGCAACCATCAGAGACATTGCAACAATTAAAGATCAC 534

Qy  116 ProLysGlyGlyLeuProValLeuLeuLeuLeuGlyHisSerAlaAspLeuArgProGlyGlu 135
        ||| |||
Db  535 CCTAAAGGAAAAATTACCCGTATTGTTACTGGGGGACACTCGCTGATCTAAGACAGGAA 594

Qy  136 PheValAlaIleAlaIleGlySerProPheAlaLeuGlnAsnThrValThrGlyIleVal 155
        ||| |||
Db  595 TTTGTAGTGCCAATTGGAAGCCCATTTGCCCTACAGAACACTGTGACAGGGTATTGTC 654

Qy  156 SerThrArgAlaGlnArgGluGluLeuGlyLeuArgAspSerAspMetAspTyrIle 175
        ::|||
Db  655 AGCACTCGTGCTCAGCGAGATGGCAAAGAACTTGGCCTTCGGGACTCCGGATAGCCATT 714

Qy  176 GlnThrArgAspAlaIleIleAsnTyrGlyAsnSerGlyTyrProLeuValAsnLeuAspGly 195
        ||| |||
Db  715 CAGACAGAAGATGCTATTATCAATTATGGCAACTCTCAGTTAATCTGATGGT 774
```

```
RESULT 11

LOCUS       BX929862               2176 bp    mRNA    linear   VRT 02-FEB-2004
DEFINITION  Gallus gallus finished cDNA, clone ChEST656k24.
ACCESSION   BX929862
VERSION     BX929862.1  GI:41630390
KEYWORDS    .
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2176)
  AUTHORS   Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,
            Croning,M.D.R., Davies,R.M., Francis,M.D., Grafham,D.V.,
            Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,
            Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,
            Tickle,C. and Wilson,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: chickest@bms.umist.ac.uk

COMMENT     BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
            sequencing project.
            This sequence is from the
            BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
            from a library constructed by Elizabeth Bosch.
            from RNA extracted from limbs, normalised, and poly A-trimmed.
            EcoRI-NotI cut cDNA was then ligated into the vector. Vector:
            pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
            coli DH10B.

FEATURES
     source   Location/Qualifiers
              1..2176
              /organism="Gallus gallus"
              /mol_type="mRNA"
              /strain="White Leghorn, Hisex"
              /db_xref="taxon:9031"
              /clone="ChEST656k24"
              /clone_lib="CSEOQCHN38"
              /dev_stage="stage 36"

ORIGIN

Alignment Scores:
  Pred. No.:
  Score:                    9.2e-115      Length:    2176
                            1450.50       Matches:    291
```

```
Qy  196 GluValIleIleGlyIleAsnThrLeuLysValThrAlaGlyIleLeuSerPheAlaIleProSer 215
        ||| |||
Db  775 GAAGTGATTGGAATTAACACCTTGAAGGTTACAGCTGGAATTCTTTCCTTTGCTATTCCGTCG 834

Qy  216 AspArgIleThrArgPheLeuThrGluPheGlnAsnGlyIleLysAspTrpLysLys 235
        ::|||
Db  835 GACCGGCATTACTCAGTTCCTCACGAGTGCGCTTGACAAACAAAATAAAGATAGCAAAAG 894

Qy  236 ArgPheLeuGlyIleLeuArgMetArgThrIleThrProSerLeuValAspGluGluLysAla 255
        ||| |||
Db  895 CGTTTCATTGGCATCCTGAATGCTGACAATAACACCTCGCCTTGGGTGGAAGAATTGAAACAC 954

Qy  256 SerAsnSerProAspPheProGluValSerSerGlyIleTyrIleSerGlyIleTyrValGlnGluValAlaProAsn 275
        ::|||
Db  955 AATAAATGCTGATTTTCCTGAAGTGTCAGCAGGTATTTACGTACATGAAGTTGTTCCAAAT 1014

Qy  276 SerProSerGlnArgGlyIleGlyIleGlnAsnGlyGlyAspSerIleAlaProGluValValMet 295
        ||| |||
Db 1015 TCACCTTCTCATGAGAGGTATCCAAGATGGAGATAATCATCGGTTAAAGTCAATGGGGT 1074

Qy  296 ProLeuValAspSerSerGluGluGlnLeuGluValAlaLeuThrGluSerProLeuLeuLeu 315
        ||| :::
Db 1075 CCTTTGATGACTTCCAGTGACCTTCAAGAGGCTGTGATGAACGAACGTCACCTCTACTACTT 1134

Qy  316 GluValArgArgGlyAsnAspLeuLeuPheSerIleAlaProGluGluValAlaVal 334
        ||| :::
Db 1135 GAAGTTCGGAGAGGAAAAATGACTTACTTAATTTAACATTGAGGCTGAAATTGTTCATG 1191
```

```
DB:
Query Match:
Best Local Similarity:
Percent Similarity:    91.15%          Conservative:  18
                       85.84%          Mismatches:    21
                       86.39%          Indels:        9
                                       Gaps:          1

US-10-617-443B-2 (1-334) x BX929862 (1-2176)

Qy      5 LeuProAlaSerAla------------------GlyLeuHisGlnLeuSer 15
          |||||||||||                      ||||||||||||
Db    329 CTGCCCGCGTCGGCTCAGGTGCAGAAAGGAGCCCTGGGGTCCACTGAGTTCCAGC 388

Qy     16 SerProArgTyrLysPheAsnPheIleAlaAspValValGluLysIleAlaProAlaVal 35
          ||||||||||||||||:::|||||||||:|||:::|||||||:||:||||||||||
Db    389 AGTCCAAGATACAAAATTCAACTTCATAGCACGAGTGGTGGAAAAGATTGCACCTGCAGTT 448

Qy     36 ValHisIleGluLeuPheLeuArgHisProLeuProGlyArgAsnValProLeuSerSer 55
          |||||||||||:::|||||||||
Db    449 GTGCACATTGAACTTTTCCTCAGGCACCCTCTGCCTGGTCGAAATGTCCCACTTTCCAGT 508

Qy     56 GlySerGlyPheIleMetSerGlyValaGlyTyrLeuIleIleThrAsnAlaHisValSer 75
          :::|||||:||||||:||||||:||||||||||
Db    509 GGATCTGGGTTTATTATGTCTGGATTCTGGTTTTATAATGTGACCAATGCACACGGTGTGTCA 568

Qy     76 SerAsnSerAlaIleAlaProGlyArgGlnGlnLeuLeuValGlnGlnLeuAsnGlyAspSer 95
          ||| :::||||
Db    569 AGTACCAATGCTATATCCAGGGAGAGCAACAGTTGAAAGTCCAGCTACAGATAACAGGAGATACA 628

Qy     96 TyrGluValaThrIleLysAspIleLysSerAspAspIleAlaThrLeuIleIleHis 115
          |||||:||||||||||||||:|||||||:
Db    629 TATGAAGCACCATCAGAGACATTGACAGTTGCAACAATGAAGATCCAC 688

Qy    116 ProLysGlyLysLeuProValLeuLeuGlyHisSerAlaAspLeuArgProGlyGlu 135
          :::|||||||||||||||||||||||||||:::||||
Db    689 CCTAAGAAAAAAATTACCCGTATTGTTACTCGGGGACACTCTCCTGACTCTAAGACCCAGAGAA 748

Qy    136 PheValAlaIleLeuGlySerProPheAlaLeuGlnIleAsnThrValThrGlyIleVal 155
          |||||||:||||||||||||
Db    749 TTTGTAGTGGCCAATTGGGAAGCCCATTGCCCTACAGAACACTGTAACAGGTATTGTC 808

Qy    156 SerThrAlaGlnArgGlyValGlyValArgGluGlyTyrLeuArgAspSerAspMetAspTyrIle 175
          |||||||||:::||||||||||||||||||||||||||||||
Db    809 AGCACTGCTCAGCGAGATGGCAAAGAACTTGGCTTCGGATATGGATTACATT 868

Qy    176 GlnThrAspAlaIleLeuGluAsnTyrGlyAsnSerGlyProLeuValAsnLeuAspGly 195
          :::|||||||||||||||||||||||||:::||||
Db    869 CAGACAGATGCTATTATCAATTATATGGCAACTCTGGAGGACCCTCTAGTTAATTCTGGATGGT 928

Qy    196 GluValIleGlyIleAlaAsnThrLeuGlyIleSerPheAlaIleProSer 215
          |||||||||||||||::||||||||||||||||||||:||||||
Db    929 GAAGTGATTGGAATTAACACCTTGAAGGTTACAGCTGAATTTCCTTTGTCATTCCGTCG 988

Qy    216 AspArgIleThrArgPheLeuThrGluProGluLysIleLeuLysAspTrpLysLys 235
          :::||||:||||
Db    989 GACCCGATTACTCAGTTCCTCACAGAGTCCGCTTGACAAACAAAATAAAGATAAGCAAAAAG 1048

Qy    236 ArgPheIleGlyIleAlaArgMetArgThrIleThrProSerLeuValAlaSpGluLeuLysAla 255
          |||||||||||||||||||||||||||||||||||||::|||||||:::||||||
Db   1049 CGTTTCATTGGCATCGCAAGGATGCAGACAATAACCGTTGGTGGAGAATTGAAACAC 1108

Qy    256 SerAsnProAspPheProGlyLeuValSerGlyLyleTyrValGlnGluValAlaProAsn 275
          |||||||||||||||||||
Db   1109 AATAATGCTGATTTCCCTGATGTCAGGAGTGGAATTTACGTACAGGAAGTTGTTCCAAAT 1168

Qy    276 SerProSerGlnArgGlyGlyIleGlyGlnAspGlyAspIleIleGluValLysValAsnGlyArg 295
          |||||||||||||||:::|||||||||||
Db   1169 TCACCTTCTCATAGAGGAGGTATCCAAGATGGTGATATCATCGTTGAAGTCAAATGGCGT 1228

Qy    296 ProLeuValAspSerSerGluLeuGlnGluValAlaValLeuThrGluSerProLeuLeuLeu 315
          |||||||:::|||||||||||||||||||
Db   1229 CCTTTGATGACTTCCAAGAGGCTGTGATGACGAATCACCTCTACTACTT 1288

Qy    316 GluValArgArgGlyValAsnAspAspLeuPheSerIleAlaLeuProGluValValMet 334
          |||||||||||||||||||||::||||
Db   1289 GAAGTTCGGAGAGGAAAATGATGACTTACTATTTAACATTGAGCCCTGAAATTGTCATG 1345
```

```
Db  1289 GAAGTTCGGAGAGGAAAATGATGACTTACTATTTAACATTGAGCCCTGAAATTGTCATG 1345

Db  1289    2080 bp    mRNA    linear  VRT 16-AUG-2004
LOCUS       BC078402  Danio rerio zgc:91963, mRNA (cDNA clone MGC:91963 IMAGE:7041654),
DEFINITION  Danio rerio zgc:91963, mRNA (cDNA clone MGC:91963 IMAGE:7041654),
            complete cds.
ACCESSION   BC078402
VERSION     BC078402.1  GI:50418456
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 2080)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodriguez,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2080)
  AUTHORS   Direct Submission
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Len Zon, Harvard
            cDNA Library Preparation: Open Biosystems
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 178 Row: 1 Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..2080
```

ORIGIN

```
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:91963 IMAGE:7041654"
/tissue_type="Whole, adult zebrafish"
/clone_lib="NIH_ZGC_10"
/lab_host="DH10B"
/note="Vector: pExpress1"
```

gene
```
1. .2080
/gene="zgc:91963"
/db_xref="LocusID:445108"
```

CDS
```
143. .1612
/gene="zgc:91963"
/codon_start=1
/product="zgc:91963"
/protein_id="AAH78402.1"
/db_xref="GI:50418457"
/db_xref="LocusID:445108"
/translation="MQVILLAAATFIPTQDIAIAELKSNCPSKCDVSKCPSPSCPSGY
VPDRCNCCLVCSLGEGAPCGRKEDVLCGDGLECKPPTGKRLSKGYVCQCKTIHRVGGTD
GKTYGNVCKMRTASRKAQQRRLSVITQRHKGSCAPSSSDPAHLNSPRYKFNIADVVE
KIAPAVVHVELFLNHPLFGRHVPLSSSGFIMTQSGLIVTNAHVVASSAVTVGRGHLR
VQLHDGQTYEASIRDIDKKSDIATIKINPKKKLQVLSIGRSADLRPGEFVVAIGSPFA
LQNTVTTGIVSTTQRDGKELGIRDSDMGYIQTDAIINYGNSSGPIVNLDGEVIGINTL
KVTAGISFAIPSDRINKFLDESNDKQQKVKQRVVRTNYTQSQAMRTASDVNVPMKRFI
GIKMVLTENLVHELKWHNPAFPDIGSGILVHEVIADSPAQKGGLESSGDIIVKLNGHP
LMNTGELQEBAIQVDMPLLLEVRGNDDLLPNIEPQILMQ"
```

Alignment Scores:

Alignment Scores:		
Pred. No.:	1. .1366-96	
Score:	1238.50	
Percent Similarity:	81.82%	
Best Local Similarity:	69.60%	
Query Match:	73.76%	
DB:	5	

	Length:	2080
	Matches:	245
	Conservative:	43
	Mismatches:	41
	Indels:	23
	Gaps:	1

US-10-617-443B-2 (1-334) x BC078402 (1-2080)

```
Qy   6 ProAlaSerAlaLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAla 25
         |||::|||||||::|::|||
Db   551 CCCTCAAGCTCAGACCTGCGCCATCTGAACAGCCCAGGTTACAAGTTTAACTTCATCGCT 610
```

```
Qy  26 AspValValGlyLysIleAlaProAlaValAlaValHisIleGluLeuPheLeuArgHisPro 45
         ||:||||||||||::|||
Db  611 GATGTGGTTGGAGAAGATCGCCCCTGCTGTGGTTCCACGTCGAGCTCTTCTTCAACCATCCA 670
```

```
Qy  46 LeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGly 65
         |||::|||::|||||:|||
Db  671 CTCTTTGGCCGTCACGTACCTTTGTCAAGCGGCGCTCTGGATTCATCATGACCGAGTCTGGT 730
```

```
Qy  66 LeuIleIleThrAsnAlaHisValValAlaSerSerAsnAlaAlaProGlyArgGlnGln 85
         ||:|||||||:||||||||
Db  731 CTGGATCGTAACCGCTCATGTGGTTGCTACAGTGGTTGCTACCAGTAACAGCAGCCGCCAGCAT 790
```

```
Qy  86 LeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLys 105
         |::|:|||||||||:::|||
Db  791 CTACGAGTGCAGCTACACGATAGCCAAACCTACGAGAGCCCTTCACGAGCAATTGACAAG 850
```

```
Qy 106 LysSerAspIleAlaThrIleLysIleHisProLysGlyLysLysProValLeuLeuLeu 125
         ||:|||||||||::||::|||
Db 851 AAATCAGATCGCCACGATCAAGATTAACCAAAGAAGAAAATACAGGTCTATCATTA 910
```

```
Qy 126 GlyHisSerAlaAspLeuArgProGlyGluPheValValAlaIleGlySerProPheAla 145
         ||||||||||:|::||||||
Db 911 GGTCGCTCTGCTGACCTGAGGCCTGGTGAGTTTGTGGTTGCCATCGGATCGCCCTTTGCT 970
```

```
Qy 146 LeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnThrAspGlyArgGluLeu 165
         |||::|||||::|||||:|::|||
Db 971 CTTCAGAACACAGTCACAACAGGCATCGTCAGCACAGCACAGACGGATGGAAGAGCTC 1030
```

```
Qy 166 GlyIleArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsn 185
         |||:::|||||||||||||
Db          (continues)
```

RESULT 13

```
LOCUS       AY280666                1953 bp    mRNA    linear   PRI 23-MAY-2003
DEFINITION  Homo sapiens pregnancy-related serine protease HTRA3 mRNA, complete
            cds; alternatively spliced.
ACCESSION   AY280666
VERSION     AY280666.1  GI:31044219
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1953)
  AUTHORS   Nie,G.Y., Hampton,A., Li,Y., Findlay,J.K. and Salamonsen,L.A.
  TITLE     Identification and cloning of two isoforms of human
            high-temperature requirement factor A3 (HtrA3), characterization of
            its genomic structure and comparison of its tissue distribution
            with HtrA1 and HtrA2
  JOURNAL   Biochem. J. 371 (Pt 1), 39-48 (2003)
  MEDLINE   22533278
   PUBMED   12513693
REFERENCE   2  (bases 1 to 1953)
  AUTHORS   Nie,G.Y., Hampton,A., Li,Y., Findlay,J.K. and Salamonsen,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-2003) Prince Henry's Institute of Medical
            Research, 246 Clayton Rd., Melbourne, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1953
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4p16.1"
                     /tissue_type="heart"
     CDS             108. .1181
```

```
Db 1031 GGTATCCCGGGACTTGACATGGGGCTACATCCAGACCGATGCCATCATCAATTATGGAAAC 1090
           ||||||||||||||||||||||
Qy  186 SerGlyGlyTyrProLeuValAsnLeuAspTyrGlyGluValIleGlyIleAsnThrLeuLysVal 205
```

```
Db 1091 TCAGGAGGTTCCACTGGTTAATCTGGATGATGGTGAGGTTGAGGGTGAATAAAACACATTAAAAGTG 1150
           ||||||||||||:||:||||||
Qy  206 ThrAlaGlyIleSerProPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPhe 225
```

```
Db 1151 ACTGCAGGGAGTCTCTTTGCCATTCCGTCAGACCGAATCAACAAGTTCCTTGATGAATCT 1210
           ||:||||||||||||||:||||
Qy  226 GlnAspLysGlnIleLysAspTrpLys------------------------------ 234
```

```
Db 1211 AATGACAAAACAGCAGAAGTCAAGCAAAGGGTAGTCAGGACAAACTACACGGCAGTCTCAA 1270
                    :::
Qy  235 -----------------------------------LysArgPheIleGlyIleArgMet 242
```

```
Db 1271 GCTATGAGGACTGCCTCAGACGTAAATGTCCCATGAGAAAGGTTTATTGGAATCAAAATG 1330
           ||||||||||||::||:|::
Qy  243 ArgThrIleThrProSerLeuValAspGluLeuLysAspArgPhePheProGlu 262
```

```
Db 1331 GTCACTTTAACAGAGAACCTAGTCCATGAATTGAAATGGCATAATCCAGCCTTTCCTGAT 1390
           |||||:||||:||||:||
Qy  263 ValSerSerGlyIleLeuTyrValGlnGluValAlaProAsnSerProSerGlnArgGlyGly 282
```

```
Db 1391 ATTGGAGTGGGGATTCTTGTTCATGAAGTTATCGCAGACTCCCCTGCTCAGAAGGGTGGG 1450
           :::|||||||||||||||:||
Qy  283 IleGlnAspGlyAspIleIleIleLysValGlnAsnGlyArgProLeuValAspSerSerGlu 302
```

```
Db 1451 CTTGAAATCAGGTGAATATATTAGTGAAACTAAACGGGCCATCCATTAATGAACACTGGAGAA 1510
           :::::
Qy  303 LeuGlnGluAlaValLeuThrGluSerProLeuLeuLeuLeuGluValArgGlyAsnAsp 322
```

```
Db 1511 CTGCAGGAAGCCGATTCAGGTGGCACATGCCCCTCCCTTTTGGAGGTTCGCGCGTAATGAC 1570
           ||:||||||||||:::|||
Qy  323 AspLeuLeuProAsnIleAlaProGluGluValValValMet 334
```

```
Db 1571 GATCTGCTTTTTAACATTGAACCTCAAATCCTCATG 1606
           ||||:||
```

```
RESULT 14
```

```
                /note="alternatively spliced"
                /codon_start=1
                /product="pregnancy-related serine protease HTRA3"
                /protein_id="AAP42283.1"
                /db_xref="GI:31044220"
                /translation="MQARLLLAALALAALLAREPPAAPCPARCDVSRCPSPRCPGGIV
                PDLCNCCLVCAAASEGEPCGGPLPDSPCGESLBCRVRGLCRCRWSHAVCGTDGHTYANVCA
                LQAASRRALQLSGTPVRQLQKGACPLGLHQLSSPRYKFNPIADVVEKIAPAVVHIELF
                LRHPLFGRNVPLSSGSGFIMSEAGLIITNAHVVSSNSAPGQROLKVQLQNGDSYEAT
                IKDIDKKSDIATIKIHPKKKLPVLLLGHSADLRPGEFVVAIGSPFALQNTVTTGIVST
                AQRGRELGLRDSDMDYIQTDAIINYGNSGGPLVNLDGEVIIGINTLKVTAGISFAIPS
                DRITRFLTEFQDKQIKAPSLAVH"
ORIGIN
```

US-10-617-443B-2 (1-334) x AY280666 (1-1953)

```
Alignment Scores:
Pred. No.:              10
Score:               3.34e-86        Length:       1953
Percent Similarity:  1117.00         Matches:       222
                     100.00%         Conservative:    0
Best Local Similarity: 100.00%       Mismatches:      0
Query Match:          66.53%         Indels:          0
DB:                    9             Gaps:            0
```

```
Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrTyrLysPheAsnPheIleAlaAspValValGlu 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492 GGTCTCCACCAGCTGAGCAGCCGCGCCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 551

Qy   30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  552 AAGATCGCACCAGCCGTGGTGCATATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 611

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  612 AACGTGCCCCTGTCCAGCGGTTCTGGCCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 671

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnLeuLysValGln 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  672 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCTCCCCCGGGCAGGCAGCTCAAGGTGCAG 731

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAspIle 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  732 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCCGACATT 791

Qy  110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  792 GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCC 851

Qy  130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  852 GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 911

Qy  150 ValThrThrGlyIleValSerThrAlaGlnArgGlyArgGluLeuGlyLeuArgAsp 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  912 GTGACAACGGGCATGGTCAGCACTGCCCAGCGGGGCCGGGAGCTGGGCCTCCGGGAC 971

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  972 TCCGACATGGACTACATCCAGACAGATGCCATCATCAACTACGGGAACTCCGGGGGACCA 1031

Qy  190 LeuValAsnLeuAspGlyGluValIleIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1032 CTGGTGAACCTGGATGGCGAGGTCATCATCGGCATCAACACGCTCAAGGTCACGGCTGGCATC 1091

Qy  210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1092 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 1151

Qy  230 IleLys 231
         ||||||
Db 1152 ATCAAA 1157
```

```
CQ716317
LOCUS       CQ716317    819 bp   DNA    linear   PAT  03-FEB-2004
DEFINITION  Sequence 2251 from Patent WO02068579.
ACCESSION   CQ716317
VERSION     CQ716317.1  GI:42277174
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 2251 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..819
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

US-10-617-443B-2 (1-334) x CQ716317 (1-819)

```
Alignment Scores:
Pred. No.:              10
Score:               1.91e-85        Length:        819
Percent Similarity:  1102.50         Matches:       235
                      73.23%         Conservative:    3
Best Local Similarity: 72.31%        Mismatches:     10
Query Match:          65.66%         Indels:         77
DB:                    6             Gaps:            3
```

```
Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrTyrLysPheAsnPheIleAlaAspValValGlu 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 GGTCTCCACCAGCTGAGCAGCCGCGCCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 132

Qy   30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133 AAGATCGCACCAGCCGTGGTGCATATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 192

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193 AACGTGCCCCTGTCCAGCGGTTCTGGCCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 252

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnLeuLysValGln 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCTCCCCCGGGCAGGCAGCTCAAGGTGCAG 312

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAspIle 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCCGACATT 372

Qy  110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  373 GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCC 432

Qy  130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  433 GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 492

Qy  150 ValThrThrGlyIleValSerThrAlaGlnArgGlyArgGluLeuGlyLeuArgAsp 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  493 GTGACAACGGGCATGGTCAGCACTGCCCAGCGGGGCCGGGAGCTGGGCCTCCGGGAC 552

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  553 TCCGACATGGACTACATCCAGACAGATGCCATCATCAACTACGGGAACTCCGGGGGA-------- 597

Qy  190 LeuValAsnLeuAspGlyGluValIleIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
Db  597 ------------------------------------------------------------- 597
```

```
Qy  210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
                                        |||
Db  598 -------CCCAGGGACAGGAGGCGGGGGCGGCACCTG------------------------ 624

Qy  230 IleLysAspTrpLysLysArgPheIleGlyIleLeuArgMetArgThrIleThrProSerLeu 249
                                                          :::
Db  625 ------------------------------------------CCATCTGTG 633

Qy  250 ValAspGluGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
Db  633 ---------------------------------------- 633

Qy  270 GlnGluValAlaIleAlaProAsnSerProSerGlnArgGlyIleGlnIleIle 289
               :::
Db  634 ------CCCAAGACCTCAAACCTCGACAGAGGCGGCCATCCAAGATGGTGACATCATC 681

Qy  290 ValLysValAlaAsnGlyAlaArgProLeuValAlaAspSerSerGluLeuGlnGluValAlaValLeuThr 309
        |||                                         |||
Db  682 GTCAAGGTCAACGGCGCGTCCTCTAGTGGACTCGAGTGAGCCTGCAGGAGCCGTGCCTGACC 741

Qy  310 GluSerProLeuLeuGluGluGlnValAlaIleArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
        |||||||||||||
Db  742 GAGTCTCCTCCTCCTTACTGAGGTCGGCGGGGGGAGCGGGGG 801

Qy  330 ProGluValValMet 334
        ||||||||||||
Db  802 CCTGAGGTGGTCATG 816
```

RESULT 15

LOCUS   AY280664   1897 bp   mRNA   linear   ROD 01-MAY-2004
DEFINITION Mus musculus pregnancy-related serine protease mRNA, complete cds;
            alternatively spliced.
ACCESSION   AY280664
VERSION   AY280664.1 GI:33358216
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1897)
AUTHORS   Nie,G., Li,Y., Findlay,J.K. and Salamonsen,L.A.
TITLE   Identification and cloning of a novel pregnancy-related
            serine-protease (PRSP)
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1897)
AUTHORS   Nie,G., Li,Y., Findlay,J.K. and Salamonsen,L.A.
TITLE   Direct Submission
JOURNAL   Submitted (22-APR-2003) Prince Henry's Institute of Medical
            Research, 246 Clayton Rd, Melbourne, Victoria 3168, Australia
FEATURES   Location/Qualifiers
     source   1..1897
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Swiss"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /tissue_type="uterus"
     CDS   127..1218
                     /codon_start=1
                     /note="alternatively spliced"
                     /product="pregnancy-related serine protease"
                     /protein_id="AAQ16583.1"
                     /db_xref="GI:33358217"
                     /translation="MQARLLPATLAILATLLAVILALAREPPAAPCPARCDVSRCPSPR
                     CPGGYVPDLCNCLVCAASEGEPCGRPLDSPGCDSLECVRGVCRCRWTHTVCGTDGHT
                     YADVCALQAASRRALQVSCTPVRQLQKGRACPSGLHQLTSPRYKPNFIADVVEKIAPAV
                     VHIELFLRHPLFGRNVPLSSGSGFIMSEAGLIVTNAHVVSSSSTASGRQQLKVQLQNG
                     DAYEATIQDIDKKSDIATIVIHPKKLPVLLFGHSADLRPGERVVAIGSPPALQNTVT
                     TGIVSTAQRDGKELGLRDSDMDYIQTDAIINYGNSGGPLVNLDGEVIGINTLKVAAGI
                     SFAIPSDRITRPLSEFQNKHVKALSPALH"
ORIGIN

Alignment Scores:
Pred. No.:   4.53e-81   Length: 1897
Score: 1057.00   Matches: 207
Percent Similarity: 97.76%   Conservative: 11
Best Local Similarity: 92.83%   Mismatches: 5
Query Match: 62.95%   Indels: 0
DB: Gaps: 0

US-10-617-443B-2 (1-334) x AY280664 (1-1897)

```
Qy    9 AlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValVal 28
        :::||||||||||||||:||::||||||||||||||||||||||||
Db  526 TCTGGTCTCCACCAGCTGACCAGTCCGCGGGGTACAAGTTCAACTTCATCGCCGATGTGGTG 585

Qy   29 GluLysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGly 48
        |||||||||||||||:||||||||||||||:|||||:||:||||||||||
Db  586 GAGAAGATTGCGCCAGCTGTTGGTCCACATAGAGCTCTTTCTGAGACACCCCTTGTTTGGC 645

Qy   49 ArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIle 68
        ::|||||||||||||||||||||||||||||||||||::|
Db  646 CGGAATGTGCCGCTGTCCTCGGGCTCGGGGTTCATCATGTCGGAAGCCGGGTTGATCGTC 705

Qy   69 ThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysVal 88
        ||||||||||||:|||||:||||||||||||||||||||||||||||||
Db  706 ACCAACGCCCACGTGGTCTCCAGCTCCAGCACTGCCTCCGGCCAGCAGCTGAAGGTG 765

Qy   89 GlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAsp 108
        ||||||||||||:||::|||||||||||||||||:||:||:|||||||
Db  766 CAGCTGCAGAATGGGGATAGCTATGAGGCCACCATCCAGGACATCGACAAGAAGTCGGAC 825

Qy  109 IleAlaThrIleLysHisProLeuValGlyLysValLeuLeuLeuGlyHisSer 128
        |||||:|||:|||||||||::||::||
Db  826 ATTGCCACGATTGTAATCCACCCAAGAAAAAAGCTCCCTGTTGTTCTGCTGGGTCACTCA 885

Qy  129 AlaAspLeuArgProGlyGluPheValValAlaIleGlySerProProAlaLeuGlnAsn 148
        ||||||||||::||||:||||||||||||||||||::||
Db  886 GCAGACCTGCCGCCCTGGGCAGTTCGTGCGTGGCCATCGGGAGCCCGCCCGCTTTGCCGAAC 945

Qy  149 ThrValThrThrGlyIleValSerThrAlaGlnArgAspGlyLysGluLeuGlyLeuArg 168
        :|||||||||||||||||||:||||||||||||||||||||||||
Db  946 ACCGTGACAACGGGCATTGTCAGCACTGCCCAGCGGGATGGGAAGGAGCTGGGTCTCCGG 1005

Qy  169 AspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGly 188
        :|||||||||:||||||||||||||||:||::||||||||||||
Db 1006 GACTCAGACATGGACTATATCCAGACCGATGCCATCATCAATTACGGGAACTCAGGAGGA 1065

Qy  189 ProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValAlaThrGly 208
        ||||||||||||||:|||||||||||:||:||||||||||:||||:|||
Db 1066 CCCCTGGTGAACCTGGATGGCGAGGTCATCGGGCATCAACACGGTTCAAGGTTGCAGCTGGC 1125

Qy  209 IleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLys 228
        |||:|||||||||||||||||:||||||||||||||::|||::|||
Db 1126 ATCTCCTTTGCCATCCCTTCAGATCGCATCACAAGGCTTCCTCTCGAGTTCCAAAACAAG 1185

Qy  229 GlnIleLys 231
        :::||||
Db 1186 CATGTGAAA 1194
```

Search completed: June 29, 2005, 23:48:50
Job time : 7535.54 secs

```
                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      June 29, 2005, 18:37:40 ; Search time 937.732 Seconds
                    (without alignments)
                    2108.485 Million cell updates/sec

Title:          US-10-617-443B-2
Perfect score:  1679
Sequence:       1 MHLALPASAGLHQLSSPRYK.........LEVRRGNDDLLFSIAPEVVM 334

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop 6.0 , Delext 7.0

Searched:       4390206 seqs, 295987066 residues

Total number of hits satisfying chosen parameters:   8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10617443//runat_29062005_135931_26052/app_query.fasta_1.718
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOPCL=0.
-LOOPEXT=0 -UNITS=bits -START=-1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LOPEXT=0 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-LIST=45 -DOCALIGN=200 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-MODE=LOCAL -OUTFMT=pto -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10617443 @CGN 1.1.586 @runat_29062005 135931 26052 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   1:  N_Geneseq_16Dec04:*
             2:  geneseqn1980s:*
             3:  geneseqn1990s:*
             4:  geneseqn2000s:*
             5:  geneseqn2001as:*
             6:  geneseqn2001bs:*
             7:  geneseqn2002as:*
             8:  geneseqn2002bs:*
             9:  geneseqn2003as:*
             10: geneseqn2003bs:*
             11: geneseqn2003cs:*
             12: geneseqn2003ds:*
             13: geneseqn2004as:*
                 geneseqn2004bs:*
```

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	1679	100.0	3006	12	ADJ11355	Adj11355 Human PRS
2	1635	97.4	1044	6	AAD44189	Aad44189 Mouse TAN
3	1635	97.4	1316	12	ADO01101	Ado01101 Human ser
4	1635	97.4	1319	12	ADO01102	Ado01102 Human ser
5	1635	97.4	1359	12	ADO01091	Ado01091 Human ser
6	1635	97.4	1375	12	ADO01099	Ado01099 Human ser
7	1635	97.4	1378	12	ADO01093	Ado01093 Human ser
8	1635	97.4	1563	3	AAA57361	Aaa57361 cDNA enco
9	1635	97.4	1583	10	ACF16993	Acf16993 Human NOV
10	1635	97.4	1797	6	AAD23855	Aad23855 Human pro
11	1635	97.4	2040	3	AAZ52362	Aaz52362 NSEQ gene
12	1635	97.4	2094	6	AAL42458	Aal42458 Human mat
13	1635	97.4	2543	8	ABK51497	Abk51497 cDNA enco
14	1635	97.4	2554	12	ACF12810	Acf12810 Human PRS
15	1635	97.4	2577	3	ADO01092	Ado01092 Human ser
16	1635	97.4	2450	8	AAA57359	Aaa57359 cDNA enco
17	1632	97.2	1375	12	ADO01106	Ado01106 Human ser
18	1632	97.2	1378	12	ADO01105	Ado01105 Human ser
19	1608	95.5	960	12	ADO01116	Ado01116 Human ser
20	1536	91.5	1075	12	ACF12809	Acf12809 Full leng
21	1524	90.8	1338	13	ADT49833	Adt49833 Murine TA
22	1378	82.1	1868	4	AAS41158	Aas41158 cDNA enco
23	1378	82.1	1868	4	AAS26920	Aas26920 Human cDN
24	1144	68.1	1075	12	ADO01100	Ado01100 Human ser
25	1122	66.8	1695	4	AAS14885	Aas14885 Human cDN
26	1117	66.5	1953	8	ACF12811	Acf12811 Human PRS
27	1105	65.8	1788	8	AAS26848	Aas26848 Human cDN
28	1057	63.0	1897	8	ACF12815	Acf12815 Full leng
29	1034.5	61.6	1185	3	AAZ29195	Aaz29195 cDNA enco
30	1034.5	61.6	1440	6	AAD44188	Aad44188 Human TAN
31	1034.5	61.6	1443	10	AAZ29176	Aaz29176 Recombina
32	1034.5	61.6	1443	10	ADL13864	Adl13864 Osteoarth
33	1034.5	61.6	1443	10	ADL13865	Adl13865 Osteoarth
34	1034.5	61.6	1695	4	ADB62641	Adb62641 Human cDN
35	1034.5	61.6	1973	13	ADL13863	Adl13863 Osteoarth
36	1034.5	61.6	2034	7	ADS10059	Ads10059 Human the
37	1034.5	61.6	2036	13	ADS73135	Ads73135 Human kid
38	1034.5	61.6	2036	13	ABV78051	Abv78051 Hypoxia-r
39	1034.5	61.6	2039	10	ADR24954	Adr24954 Breast ca
40	1034.5	61.6	2054	10	ADR40166	Adr40166 Human ser
41	1034.5	61.6	2054	10	ADL13867	Adl13867 Osteoarth
42	1034.5	61.6	2054	10	ADJ56226	Adj56226 Rat cDNA
43	1034.5	61.6	2075	10	ADL13863	Adl13863 Osteoarth
44	1034.5	61.6	2205	2	AAT75444	Aat75444 cDNA enco
45	1034.5	61.6	2205	2	AAV04680	Aav04680 Human pre
	1034.5	61.6	2437	11	ACN89676	Acn89676 Breast ca

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

```
RESULT 1
ADJ11355
ID   ADJ11355 standard; cDNA; 3006 BP.
XX
AC   ADJ11355;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Human PRSS11-L cDNA encoding a serine protease SeqID 1.
XX
KW   human; PRSS11-L; gene; ss; S2 serine protease; S2/HtrA;
KW   cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;
KW   apoptotic; osteopathic; antiarthritic; tranquiliser.
XX
OS   Homo sapiens.
XX
FH   Key         Location/Qualifiers
FT   CDS         1011..2015
FT               /*tag=  a
FT               /product= "PRSS11-L protein"
XX
PN   US2004005659-A1.
XX
PD   08-JAN-2004.
XX
PF   03-JUL-2002; 2002US-00189099.
XX
PR   03-JUL-2002; 2002US-00189099.
```

```
PA   (DARR/) DARROW A L.
PA   (QLJÜ/) QL J. J.
PA   (CHEN/) CHEN C.
PA   (ANDR/) ANDRADE-GORDON P.
XX
PI   Darrow AL, Ql J, Chen C, Andrade-Gordon P;
XX
DR   WPI; 2004-081723/08.
DR   P-PSDB; ADJ11356.
XX
PT   New isolated S2 serine protease nucleic acids and polypeptides, useful
PT   for diagnosing and/or treating diseases with aberrant expression or
PT   activity the S2 serine protease, such as osteoarthritis, stress and
PT   apoptotic disorders.
XX
CC   This invention relates to a novel isolated nucleic acid molecule
CC   identified as PRSS11-L that encodes an S2 serine protease. Specifically,
CC   it refers to members of the S2/HtrA serine protease family, such that it
CC   plays a role in cellular physiology and apoptosis. The present invention
CC   provides agonists, antagonists, antibodies and recombinant expression
CC   vectors useful in methods of treatment, or detection and diagnosis of
CC   diseases associated with the aberrant expression or activity of the S2
CC   serine protease, PRSS11-L. According, compositions described herein can
CC   be used via gene therapy routes to treat osteoarthritis, stress and
CC   apoptotic disorders. As such, they exhibit osteopathic, antiarthritic and
CC   tranquiliser activities. This polynucleotide sequence is the human PRSS11
CC   -L cDNA sequence of the invention.
XX
PS   Claim 2; SEQ ID NO 1; 28pp; English.
XX
SQ   Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:
Score:                    9.03e-164
Score:                    1679.00               Length:            3006
Percent Similarity:       100.00%               Matches:           334
Percent Identity:         100.00%               Mismatches:        0
Best Local Similarity:    100.00%               Conservative:      0
Query Match:              12                     Indels:            0
DB:                                             Gaps:              0
```

```
US-10-617-443B-2 (1-334) x ADJ11355 (1-3006)

Qy    1 MetHisLeuAlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerProArgTyrTyrLys 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1011 ATGCACCTGGCCCTTCCCGCTAGCGCAGGGCTCCCACCACGTACAGCAGCCCCGCTACAAG 1070

Qy   21 PheAsnPheIleAlaAspValValGlyIleAlaProAlaValValHisIleGluLeu 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1071 TTCAACTTCATTGCTGACGTGGTGGGAGAAGCGCCACCAGCCGTGGTCCACATAGAGCTC 1130

Qy   41 PheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1131 TTCCTGAGACACCCGCTGTTTGGCCGCAACGTGCCCCTGTCCAGCGGGTCTGGCTTCATC 1190

Qy   61 MetSerGluAlaGlyLeuIleLeuThrAsnAlaHisValSerSerAsnSerAlaAla 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1191 ATGTCAGAGGCCGGCCTGATCATCACCAATGCCCACGTGTCCAGCAACAGTGCTGCC 1250

Qy   81 ProGlyArgGlnGlnLeuLeuValGlnLeuGlyAspSerTyrGluAlaThrIle 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1251 CCCGGCCAGGCAGCTCAAGGTGCAGCTACAGAATGGGAGCTCCTATGAGGCCACCATC 1310

Qy  101 LysAspIleAspGlyLysSerAspIleAlaThrIleLysIleHisProLysLysSerLeu 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1311 AAAGACATCGACAAGAAGTCGGACATTGCCACCATCAAGATCCATCCCAAGAAAAAGCTC 1370

Qy  121 ProValLeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValValAlaIle 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1371 CCTGTGTTGCTGGGTCACTCCGCCGACCTGCGGCCTGGGGAGTTTGTGGTGGCCATC 1430

Qy  141 GlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArg 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1431 GGCAGTCCCTTCGCCCTACAGAACACAGTGACAACGGGCATCGTCAGCACTGCCCACGG 1490

Qy  161 GluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1491 GAGGGCAGGGAGCTGGGCCTCCGGGACTCCGACATGGACTACATCCAGACGGATGCCATC 1550

Qy  181 IleAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIle 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1551 ATCAACTACGGGAACTCCGGGGGCCCTCTGGTGAACCTGGATGGGGAGGTCATTGGCATC 1610

Qy  201 AsnThrLeuLysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArg 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1611 AACACGCTCAAGGTCACCGCTGGCATCTCCTTTGCCATCCCCAGCGACCGCATCACACGG 1670

Qy  221 PheLeuThrGluPheGlnAspLysIleLysLysAspTrpLysLysArgPheIleGlyIle 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1671 TTCCTCACAGAGTTCCAGGACAAGATCAAGAAAGACTGGAAGAAGCGCTTCATCGGCATA 1730

Qy  241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAspPhe 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1731 CGGATGCGGACGATCACACCAAGCCTGGTGGATGAGCTGAAGGCCAGCAACCCGGACTTC 1790

Qy  261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1791 CCAGAGGTCAGCAGTGGGATTTATGTGCAAGAGGTTGCGCCGAATTCACCTTCTCAGGA 1850

Qy  281 GlyGlyIleGlnAspGlyAspIleIleValLysValAsnGlyArgProLeuValAspSer 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1851 GGGGGCATCCAAGATGGTGACATCATCGTCAAGGTCAACGGGCGTCCTCTAGTGGACTCG 1910

Qy  301 SerGluLeuGlnAlaValLeuThrGluSerProLeuLeuLeuGluValArgArgGly 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1911 AGTGAGCTGCAGGCGGTGCTGACCGAGTCTCCTCTACTGGAGGTGCGCCGGGGG 1970

Qy  321 AsnAspAspLeuLeuPheSerIleAlaProAlaProGluValValMet 334
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1971 AACGACGACCTCCTCTTCAGCATCGCACCTGCACCTGAGGTGGTCATG 2012
```

```
RESULT 2
AAD44189
ID   AAD44189 standard; cDNA; 1044 BP.
XX
AC   AAD44189;
XX
DT   13-DEC-2002 (first entry)
XX
DE   Mouse TANGO 219 cDNA.
XX
KW   TANGO protein; vaccine; chromosomal mapping; antisense gene therapy;
KW   forensic biology; predictive medicine; cytostatic; pharmacogenomic;
KW   tissue typing; mouse; gene; ss.
XX
OS   Mus musculus.
OS   Mus musculus.
XX
PN   US200205139-A1.
XX
PD   09-MAY-2002.
XX
PF   01-MAR-2001; 2001US-00796858.
XX
PR   30-DEC-1998; 98US-00223094.
PR   30-DEC-1998; 98US-00223546.
PR   30-DEC-1998; 98US-00224246.
PR   14-MAY-1999; 99US-00312359.
PR   18-JUN-1999; 99US-00336836.
PR   29-JUN-1999; 99US-00342687.
PR   30-JUL-1999; 99US-00365104.
PR   20-SEP-1999; 99US-00399723.
PR   23-DEC-1999; 99US-00471179.
PR   29-DEC-1999; 99US-00474071.
PR   29-DEC-1999; 99US-00474072.
PR   15-MAY-2000; 2000US-00572002.
PR   19-JUN-2000; 2000US-00597993.
PR   22-JUN-2000; 2000US-00599596.
```

```
PR   29-JUN-2000; 200US-00605565.
PR   31-JUL-2000; 200US-00630334.
PR   20-SEP-2000; 200US-00665666.
XX
PA   (HOLT/)  HOLTZMAN D A.
PA   (SHAR/)  SHARP J D.
PA   (LEIB/)  LEIBY K R.
PA   (BOSS/)  BOSSONE S.
PA   (PANY/)  PAN Y.
PA   (BARN/)  BARNES T M.
PA   (FRAS/)  FRASER C C.
PA   (WRIG/)  WRIGHTON N.
PA   (MYER/)  MYERS P S.
PA   (KING/)  KINGSBURY G.
XX
PI   Holtzman DA, Sharp JD, Leiby KR, Bossone S, Pan Y, Barnes TM;
PI   Fraser CC, Wrighton N, Myers PS, Kingsbury G;
XX
DR   WPI; 2002-453953/48.
XX
PT   New isolated nucleic acid encoding a TANGO polypeptide, for use as a
PT   modulating agent for regulating cellular processes and for use in a
PT   vaccine.
XX
CC   The invention relates to TANGO polypeptide and its corresponding nucleic
CC   acid sequence. TANGO protein is used to identify a compound which binds
CC   to it that can be used for modulating its activity. It is also used to
CC   produce an antibody. The antibody is used to detect the presence of the
CC   polypeptide in a sample. TANGO DNA and protein are useful as modulating
CC   agents in regulating cellular processes. They can be used in vaccines.
CC   TANGO DNA and protein and its antibody are used in e.g. chromosomal
CC / mapping, tissue typing, forensic biology, predictive medicine,
CC   pharmacogenomics and treatment methods. TANGO DNA is used in antisense
CC   gene therapy. The present sequence is mouse TANGO 219 cDNA. Note: This
CC   sequence SEQ.ID.NO.9 is said to encode the protein AAE26441, but this is
CC   not the case
XX
PS   Claim 1; Page 81-82; 206pp; English.
XX
SQ   Sequence 1044 BP; 230 A; 318 C; 306 G; 190 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                     7.83e-160
                           1635.00
Percent Similarity:        100.00%
Percent Identity:          100.00%
Best Local Similarity:     100.00%
Query Match:               97.38%

                           Length:          1044
                           Matches:          325
                           Conservative:       0
                           Mismatches:         0
                           Indels:             0
                           Gaps:               0

US-10-617-443B-2 (1-334) x AAD44189 (1-1044)

Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 GGTCTCCACAGCTGAGCAGCCCGCCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 129

Qy   30 LysIleAlaProAlaValValHisIleGlnLeuPheLeuArgHisProLeuPheGlyArg 49
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTCGGCCGC 189

Qy   50 AsnValProLeuSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleLeuThr 69
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  190 AACGTGCCCCTGTCCGGCAGCGGTTCGGCCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 249

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyTyrArgGlnLeuLysValGln 89
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  250 AATGCCCACGTGGTCTCCAGCAACAGTGCTGCCCCGGGCTACAGGCAGCTCAAGGTGCAG 309

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysIleSerAspIle 109
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  310 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGATCTCGGACATT 369

Qy  110 AlaThrIleLeuIleHisProLysGlyLysGlyLeuProValLeuLeuLeuGlyHisSerAla 129
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  370 GCCACCATCAAGATCCATCCCAAGAAAAAAGCTCCCTGTGTTGTTGCTGGGGTCACTCGCC 429

Qy  130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  430 GACCTGCGGCCTGGGGAGTTTGTGGTTGCCATCGGAAGCCCTTTCGCCTTACAGAACACA 489

Qy  150 ValThrThrArgThrGlyIleValSerThrAlaGlnGlyLeuLeuGlyLeuLeuArgAsp 169
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  490 GTGACAACGAGAACCGGCATCGTCAGCACTGCCCAGGGACTGCTGGGCCTTCGGGAC 549

Qy  170 SerAspMetAspThrTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  550 TCCGACATGGACACATACATCCAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGACCA 609

Qy  190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  610 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC 669

Qy  210 SerPheAlaIleProSerAspAsnArgIleArgPheLeuThrArgPheGlnAspSerLysGly 229
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  670 TCCTTTGCCATCCCCTCAGACCCGCCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 729

Qy  230 IleLysAspTrpLysLysArgPheIleGlyIleGlyMetArgThrIleThrProSerLeu 249
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  730 ATCAAGACTGGAAGAAGCGCTTCATCGGGATCGGAATGCGGACGATCACACCAAGCCTG 789

Qy  250 ValAspGluLeuLysAlaSerAsnProAspPhePheGluProGluValSerSerGlyIleTyrVal 269
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  790 GTGGATGAGCTGAAGGCCAGCAATCCCGGACTTCCCAGAGGTCAGTGGAATTTATGTG 849

Qy  270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAsnGlyGlyAspIleIle 289
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  850 CAAGAGGTTGCCCCGAATTCACCTTCTCAGACAGGCCGCATCCAAAATGGTGACATCATC 909

Qy  290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  910 GTCAAGGTCAACGGGCGGCCTCTAGTGGACTCGAGTGAGCTGCAGGAGGCCGTGCTGACC 969

Qy  310 GluSerProLeuGluLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  970 GAGTCTCCTCTCCTACTGGAGGTGCGGCGGGGGAACGACGACCTCCTCTTCAGCATCGCA 1029

Qy  330 ProGluValValMet 334
         |||||||||||||||
Db 1030 CCTGAGGTGGTCATG 1044

RESULT 3
AD001101
ID   AD001101 standard; DNA; 1316 BP.
XX
AC   AD001101;
XX
DT   29-JUL-2004 (first entry)
XX
DE   Human serine protease HTRA3 related DNA, SEQ ID 16.
XX
KW   Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;
KW   serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004039407-A1.
XX
PD   13-MAY-2004.
XX
PF   30-OCT-2003; 2003WO-JP013920.
XX
PR   01-NOV-2002; 2002JP-00320075.
PR   27-JAN-2003; 2003JP-00017892.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;
```

```
XX
DR   WPI; 2004-400073/37.
XX
PT   HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.
XX
PS   Example 3; SEQ ID NO 16; 124pp; Japanese.
XX
CC   Apoptosis inducers inhibiting activity or expression of serine protease
CC   The present invention relates to apoptosis inducers and protease
CC   inhibitors, which contain a compound inhibiting the activity or
CC   expression of human serine protease HTRA3 (AD001090 and AD001115). The
CC   invention is useful for the treatment, prevention and diagnosis of
CC   pancreatic cancer, and also of other cancers including colon, breast,
CC   lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
CC   ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
CC   present sequence was used to illustrate the invention.
XX
SQ   Sequence 1316 BP; 254 A; 425 C; 404 G; 233 T; 0 U; 0 Other;
```

US-10-617-443B-2 (1-334) x AD001101 (1-1316)

```
Alignment Scores:
Pred. No.:              1.08e-159          Length:       1316
Score:                  1635.00           Matches:      325
Percent Similarity:     100.00%           Conservative: 0
Best Local Similarity:  100.00%           Mismatches:   0
Query Match:            97.38%            Indels:       0
DB:                     12                Gaps:         0
```

```
Qy   10 GlyIleHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  334 GGTCTCCACAGTTGAGCAGCCCCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 393

Qy   30 LysIleAlaProLeuAlaValHisIleLeuGluLeuPheLeuArgHisProLeuPheGlyArg 49
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  394 AAGATCGCACCACTGGCTGTCCACATAGAGCTCTTCCTTCGACACCCGTTGTTTGGCCCC 453

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  454 AACGTGCCCCTGTCCAGCGGGTTCGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC 513

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyGlnLeuLeuLysValGln 89
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  514 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCCGGGCAGCTGCTGAAGGTGCAG 573

Qy   90 LeuGlnAsnGlyAspSerTyrGluValAlaThrIleLeuAspProLysLysSerAspIle 109
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  574 CTACAGAATGGGGACTCCTATGAGGGCCACCATCAAGAAGTCGGACATT 633

Qy  110 AlaThrIleLeuSerIleHisProLysLysSerLeuProValLeuLeuGlyHisSerAla 129
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  634 GCCACCATCAAGATCCACCCCAAGAAAAAGCTCCCTGTGTTGTTGGGTCACTCGGCC 693

Qy  130 AspLeuArgProGlyPheGluAlaLeuGlySerProPheAlaLeuGlnAsnThr 149
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  694 GACCTGCGGCCTGGGGAGTTTGTGGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 753

Qy  150 ValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyLeuGlyLeuArgAsp 169
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  754 GTGACAACGGGCATCGTCAGCACTGCCCAGGGAGGGACCTGGGCCTCCGGAC 813

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaIleIleLeuAsnTyrGlyAsnSerGlyPro 189
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  814 TCCGACATGGACTACATCCAGACCGGATGCCATCATCAACTACGGGAACTCCGGGGGACCA 873

Qy  190 LeuValAsnLeuAspGlyLysIleGlyIleAsnThrLeuValThrAlaGlyIle 209
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  874 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACCCTGGTCACGGCTGGCATC 933

Qy  210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  934 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 993

Qy  230 IleLysAspTrpLysLysArgPheIleGlyArgMetArgThrIleThrProSerLeu 249
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  994 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATACGGACGATCACACCAAGCCTG 1053

Qy  250 ValAspGluLeuLysLysAlaSerAsnProAspPhePheProGluValSerSerGlyIleTyrVal 269
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1054 GTGGATGAGCTGAAGAAGGCCAGCAACCCCGACTTCCCAGAGGTCAGCAGTGGGATTTATGTG 1113

Qy  270 GlnGluValAlaProAsnSerProSerGlnArgGlyLysGlyIleGlnAspAspIleIle 289
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1114 CAAGAGGTTGCCGCGAATTCACCTTCTCGAGAGGGCAAGGGCATCCAAGATGGTGACATCATC 1173

Qy  290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluValAlaValLeuThr 309
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1174 GTCAAGGTCAACGGGCGTCCTCTTGGTGGACTCGGAGCTGGTGGTGCTGACC 1233

Qy  310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1234 GAGTCTCCTCTCCTACTGGAGGTGCGGCGCGGGAACGACGACCTTCCTTCAGCATCCA 1293

Qy  330 ProGluValValMet 334
        |||||||||||||
Db 1294 CCTGAGGTGGTCATG 1308
```

```
RESULT 4
AD001102
ID   AD001102 standard; DNA; 1319 BP.
XX
AC   AD001102;
XX
DT   29-JUL-2004 (first entry)
XX
DE   Human serine protease HTRA3 related DNA, SEQ ID 17.
XX
KW   Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;
KW   serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004039407-A1.
XX
PD   13-MAY-2004.
XX
PF   30-OCT-2003; 2003WO-JP013920.
XX
PR   01-NOV-2002; 2002JP-00320075.
PR   27-JAN-2003; 2003JP-00017892.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;
XX
DR   WPI; 2004-400073/37.
XX
PT   HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.
XX
PS   Example 3; SEQ ID NO 17; 124pp; Japanese.
XX
CC   Apoptosis inducers inhibiting activity or expression of serine protease
CC   The present invention relates to apoptosis inducers and protease
CC   inhibitors, which contain a compound inhibiting the activity or
CC   expression of human serine protease HTRA3 (AD001090 and AD001115). The
CC   invention is useful for the treatment, prevention and diagnosis of
CC   pancreatic cancer, and also of other cancers including colon, breast,
CC   lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
CC   ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
CC   present sequence was used to illustrate the invention.
XX
SQ   Sequence 1319 BP; 255 A; 425 C; 405 G; 234 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:        1.08e-159          Length:   1319
Score:            1635.00           Matches:  325
```

Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 97.38%   Indels: 0
DB: 12   Gaps: 0

US-10-617-443B-2 (1-334) x AD001102 (1-1319)

```
Qy   10 GlyLeuHisGlyIleLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu  29
         GGTCTCCACCAGCTGAGCAGCCCGCGCCCTACAAAGTTCAATTCATTGCTGACGTGGTGGAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  334 GGTCTCCACCAGCTGAGCAGCCCGCGCCCTACAAAGTTCAATTCATTGCTGACGTGGTGGAG  393

Qy   30 LysIleAlaProAlaValValHisIleGlyLeuPheLeuArgHisProLeuPheGlyArg    49
         AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  394 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC    453

Qy   50 AsnValProLeuSerSerGlySerGlyGlyTrpThrGlnAlaGlyLeuAlaIleIleLeuThr  69
         AACGTCCCCCTGTCCAGCGGTTCCGGCGGTTGGACCCAGGCCGGCCTTGCCATCATCCTCACC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  454 ...                                                              513

Qy   70 AsnAlaLeuHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln  89
         GCCACCATCAAGATCCCATCCCAAGAAAAAGCTCCCTGTTGTTGTTGGGGGGGACC...
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  514 ...                                                              573

Qy   90 LeuGlnAsnGlyAspSerTyrGlyLeuAlaThrIleIleLeuLysLysAspSerAspIle    109
         CTACAGAATGGGGACTCCTATGGGCTTGCCACCATCAAAGAACATCAAGAGAGTCGGACAT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  574 ...                                                              633

Qy  110 AlaThrIleLeuLysIleHisValValSerProLysLeuProLeuLeuLeuGlyHisSerAla 129
         GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTTGTTGTTGGGGGGGACC
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  634 ...                                                              693

Qy  130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr    149
         GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCTCCCCTTTCGCCCTACAGAACACA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  694 ...                                                              753

Qy  150 ValThrThrGlyIleValSerThrAlaGlnArgGlyGlyArgGluLeuGlyLeuArgAsp    169
         GTGACAACGGGCATCGTCAGCACTGCCCAGCGAGGCGGCCGAGAGCTGGGCCTCCGGGAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  754 ...                                                              813

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro    189
         TCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGACCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  814 ...                                                              873

Qy  190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle    209
         CTGGTGAACCTGGATGGGGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  874 ...                                                              933

Qy  210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln    229
         TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  934 ...                                                              993

Qy  230 IleLysGlyAlaGlyIleLysLysGlyGluLeuGlyIleMetArgThrLeuThrProSerLeu 249
         ATCAAAGGCTGGAGGAAAGCCGCTTCATCGGGATACCGACGATACACCACCAGCCCTG
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  994 ...                                                              1053

Qy  250 ValAspGluGluLeuLysGlnAlaSerAlaSerAsnProGluValSerSerGlyIleTyrVal 269
         GTGGATGAGGAGCTGAAGCAGGCAAGCGCCAGCAACCCCGAGGTCAGCAGTGGAATTTATGTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1054 ...                                                              1113

Qy  270 GlnGluValAlaProAsnSerProSerGlnArgGlyIleGluGlnAspGlyArgIleAla    289
         CAAGAGGTTGCGCCCAAGTTCACCTTCTCAGAGAGCGGGCATCCAAGAGTGGTGCACATC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1114 ...                                                              1173

Qy  290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr    309
         GTCAAGGTCAACGGGCGTCCTCTAGTGACTCGAGTGAGCTGCAGGAGGCCGTGCTGACC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1174 ...                                                              1233

Qy  310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla    329
         GAGTCTCCTCTACTGCTGGAGGTGCGGCGGGGAAACGACGACCTCCTCTTCTCAATCGCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1234 ...                                                              1293

Qy  330 ProGluValValMet 334
         CCTGAGGTGGTCATG
         |||||||||||||||
Db 1294 CCTGAGGTGGTCATG 1308
```

Alignment Scores:
Pred. No.:   1.12e-159   Length: 1359
Score: 1635.00   Matches: 1359
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 97.38%   Indels: 0
DB: 12   Gaps: 0

US-10-617-443B-2 (1-334) x AD001091 (1-1359)

```
Qy   10 GlyLeuHisGlyIleLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu  29
         GGTCTCCACCAGCTGAGCAGCCCGCGCCCTACAAAGTTCAATTCATTGCTGACGTGGTGGAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  385 GGTCTCCACCAGCTGAGCAGCCCGCGCCCTACAAAGTTCAATTCATTGCTGACGTGGTGGAG  444

Qy   30 LysIleAlaProAlaValValHisIleGlyLeuPheLeuArgHisProLeuPheGlyArg    49
         AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  445 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC    504
```

```
XX
OS   Homo sapiens.
XX
PN   WO200403940 7-A1.
XX
PD   13-MAY-2004.
XX
PF
XX
PD
XX
PR   01-NOV-2002; 2002JP-00320075.
PR   27-JAN-2003; 2003JP-00017892.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;
XX
DR   WPI; 2004-400073/37.
XX
PT   Apoptosis inducers inhibiting activity or expression of serine protease
PT   HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.
XX
PS   Example 3; SEQ ID NO 14; 124pp; Japanese.
XX
CC   The present invention relates to apoptosis inducers and protease
CC   inhibitors, which contain a compound inhibiting the activity or
CC   expression of human serine protease HTRA3 (AD001090 and AD001115). The
CC   invention is useful for the treatment, prevention and diagnosis of
CC   pancreatic cancer, and also of other cancers including colon, breast,
CC   lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
CC   ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
CC   present sequence was used to illustrate the invention.
XX
SQ   Sequence 1375 BP; 258 A; 447 C; 429 G; 241 T; 0 U; 0 Other;
```

```
US-10-617-443B-2 (1-334) x AD001099 (1-1375)

Alignment Scores:
Pred. No.:               1.14e-159      Length:         1375
Score:                   1635.00        Matches:        325
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             97.38%         Indels:         0
DB:                      12             Gaps:           0
```

```
Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr  69
         AACGTGCCCCTTTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACC
Db   505 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 564

Qy    70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln  89
         AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGGCGCAGCAGCTCAAGGTGCAG
Db   565 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 624

Qy    90 LeuGlnAsnGlyAspSerTyrGluAlaThrLeuTyrAlaAspIleAspLysLysSerAsp 109
         CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCGGACATT
Db   625 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 684

Qy   110 AlaThrIleLysIleHisProLysLysGlyLysIleLeuProValLeuLeuLeuGlyHisSerAla 129
         GCCACCATCAAGATCCATCCCAAGAAAAAGGCTCCCGTGTTGTTGCTGGGTCACTCGGCC
Db   685 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 744

Qy   130 AspLeuArgProGlyLeuPheValValAlaIleLeuGlySerProPheAlaLeuGlnAsnThr 149
         GACCTGCGGCCGGGCCTGGGGAGTTTGTGGTGGCCATCCTGGGCAGCCCCTTCGCCCTACAGAACACA
Db   745 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 804

Qy   150 ValThrThrThrGlyIleValSerThrAlaGlnArgGlyLysArgGluGlyLeuArgAsp 169
         GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGGCAAGCGGGAGGGCCTGCGGGAC
Db   805 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 864

Qy   170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
         TCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGACCA
Db   865 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 924

Qy   190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
         CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACACTCAAAGGTCATTGGCATC
Db   925 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 984

Qy   210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
         TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG
Db   985 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1044

Qy   230 IleLysAspTrpLeuLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
         ATCAAAGACTGGAAGAGGCGCTTCATCGGCATACGGATGCGGACGATCACACCAAGCCTG
Db  1045 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1104

Qy   250 ValAlaAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
         GTGGCAGATGAGCTGAAGGCCAGCAACCCCGACTTCCCAGAGGTCAGCAGTGGAATTTATGTG
Db  1105 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1164

Qy   270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGlyAspIleIle 289
         CAAGAGGTTGCGCCGAATTCACCTTCTCAGAGGAGGGCATCCAAGATGGTGACATCATC
Db  1165 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1224

Qy   290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
         GTCAAGGTCAACGGGCGTCCTCTAGTGGACTCGAGTGAGCTGCAGGAGGCCGTGCTGACC
Db  1225 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1284

Qy   310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
         GAGTCTCCTCTCCTGCTGGAGGTGCGGCGGGGCAACGACGACCTCCTCTTCAGCATCGCA
Db  1285 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1344

Qy   330 ProGluValValMet 334
         CCTGAGGTGGTCATG
Db  1345 |||||||||||||| 1359
```

```
RESULT 6
AD001099
ID   AD001099 standard; DNA; 1375 BP.
XX
AC   AD001099;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human serine protease HTRA3 related DNA, SEQ ID 14.
XX
KW   Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;
KW   serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.
```

```
Qy    10 GlyLeuHisGlnLeuSerSerProArgTyrLysAspPheAsnPheIleAlaAspValValGlu  29
         GGTCTCCACCAGCTGAGCAGCCCCCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG
Db   393 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 452

Qy    30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg  49
         AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTTCTGAGACACCCGCTGTTTGGCCGC
Db   453 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 512

Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr  69
         AACGTGCCCCTTTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACC
Db   513 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 572

Qy    70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln  89
         AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGGCGCCAGCAGCTCAAGGTGCAG
Db   573 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 632

Qy    90 LeuGlnAsnGlyAspSerTyrGluAlaThrLeuTyrAlaAspIleAspLysLysSerAla 109
         CTACAGAATGGGGACTCCTATGAGGCCACCACCATCAAAGACATCGACAAGAAGTCGGAC
Db   633 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 692

Qy   110 AlaThrIleLysIleHisProLysLysGlyLysIleLeuProValLeuLeuLeuGlyHisSerAla 129
         GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCCGTGTTGCTGGGTCACTCGGCCGAC
Db   693 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 752

Qy   130 AspLeuArgProGlyLeuPheValValAlaIleLeuGlySerProPheAlaLeuGlnAsnThr 149
         GACCTGCGGCCGGGCCTGGGGAGTTTGTGGTGGCCATCCTGGGCAGCCCCTTCGCCCTACAGAACACA
Db   753 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 812

Qy   150 ValThrThrThrGlyIleValSerThrAlaGlnArgGlyLysArgGluGlyLeuArgAsp 169
         GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGGCAAGCGGGAGGGCCTGCGGGAC
Db   813 ||||||||||||||||||||||||||||||||||||||||||||||||||||||| 812
```

```
Db  813  GTGACAACGGCACCATCGTCAGCAGCACTGCCCAGCGGGAGGGAGGGAGGCAGCGGAGCTGGGGCCTCCGGGAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  872

Qy  170  SerAspArgMetAspTyrIleGlnThrArgProAlaIleIleLeuAsnThrGlyAsnSerGlyGlyPro
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  189
Db  873  TCCGACATGGACTACATCCAGACGCGGGCCATCATCCTCAACACTGGCAACTCCGGGGGACCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  932

Qy  190  LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIleIle
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  209
Db  933  CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACCGCTGGCATC
         ||||||||||||||||||||||||||||||||||||||||||||||||||  992

Qy  210  SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspGlyGln
         ||||||||||||||||||||||||||||||||||||||||||||||||||  229
Db  993  TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||  1052

Qy  230  IleLeuAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu
         ||||||||||||||||||||||||||||||||||||||||||||||||||  249
Db  1053 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATCAGGATGCGGACGATCACACCAAGCCTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||  1112

Qy  250  ValAspGluGluLysLysAlaSerAsnProAspPheProGluValSerGlyIleTyrVal
         ||||||||||||||||||||||||||||||||||||||||||||||||||  269
Db  1113 GTGGATGAGGTGAAGGCCAGCAACCCGGACTTCCCAGAGGTGTCAGGCAGTGGAATTTATGTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||  1172

Qy  270  GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleLeuGlnAspGlyAspIleIle
         ||||||||||||||||||||||||||||||||||||||||||||||||||  289
Db  1173 CAAGAGGTTGCGCCCAATTCACCTTCTCAGAGAGGGGGGCATCCTGCAAGATGGTGACATCATC
         ||||||||||||||||||||||||||||||||||||||||||||||||||  1232

Qy  290  ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnAlaValLeuThr
         ||||||||||||||||||||||||||||||||||||||||||||||||  309
Db  1233 GTCAAGGTCAACGGCGGTCCTCTAGTGGACTCGAGTGAGCTGCAGGCCGTGCTGACC
         ||||||||||||||||||||||||||||||||||||||||||||||||  1292

Qy  310  GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleGluAla
         ||||||||||||||||||||||||||||||||||||||||||||||||||  329
Db  1293 GAGTCTCCCTCCTACTGGGAGGTGCGGCGGGGGAACGATGACCTCCTCTTCAGCATCGGCA
         ||||||||||||||||||||||||||||||||||||||||||||||||||  1352

Qy  330  ProGluValValMet 334
         |||||||||||||||
Db  1353 CCTGAGGTGGTCATG 1367
```

The present invention relates to apoptosis inducers and protease
inhibitors, which contain a compound inhibiting the activity or
expression of human serine protease HTRA3 (AD001090 and AD001115). The
invention is useful for the treatment, prevention and diagnosis of
pancreatic cancer, and also of other cancers including colon, breast,
lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
present sequence was used to illustrate the invention.

```
SQ  Sequence 1378 BP; 259 A; 447 C; 430 G; 242 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                      1.15e-159
Score:                      1635.00        Length:             1378
Percent Similarity:         100.00%         Matches:             325
Best Local Similarity:      100.00%         Conservative:          0
Query Match:                97.38%          Mismatches:            0
DB:                         12              Indels:                0
                                            Gaps:                  0

US-10-617-443b-2 (1-334) x AD001093 (1-1378)

Qy    10  GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu  29
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   633  GGTCTCCACCAGCTGAGCAGCCCCGCCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG  692

Qy    30  LysIleAlaProAlaValValHisIleGluLeuProLeuPheGlyArg  49
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   393  GCCACCCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTCGTTGCTGGGGTCACTCGGCC  452

Qy   110  AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuLeuGlyHisSerAla  129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   693  GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTCGTTGCTGGGGTCACTCGGCC  752

Qy   130  AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr  149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   753  GACCTGCGCCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA  812

Qy    90  LeuGlnAsnGlyAspSerTyrGluAlaThrLeuLysAspSerIleAspSerArgIle  109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   573  AATGCCCACGTGGGTGGTCCAGCGAACAGTGCTGTCCCCCGGCAGCGCAGCTCAAGG  632

Qy    70  AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln  89
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   513  AACGTGCCCGTCTCCAGCGGGTTCTGGCTTCATCATGTCAGGAGGCCCGTGATCATCACC  572

Qy    50  AsnValProLeuSerSerGlySerGlyPheHisMetSerGluValAlaGlyLeuIleLeuThr  69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   453  AAGATCGCCACCAGCCCGTGGCTCCCACATAGAGTCTTCTTCGAGACACCCGCTGTTTGGCGGC  512

Qy   150  ValThrThrArgGlyIleValSerThrArgAlaGlnArgGlyIleLeuGlyIleLysArgAsp  169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   813  GTGACAACGGCACCATCGTCAGCACTGCCCAGCGGGAGGGAGGGAGGCAGCGGAGCTGGGGGC  872

Qy   170  SerAspMetAspTyrIleGlnThrArgProAlaIleIleLeuAsnThrGlyAsnSerGlyGlyPro  189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   873  TCCGACATGGACTACATCCAGACGCGGGCCATCATCCTCAACACTGGCAACTCCGGGGGACCA  932

Qy   190  LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIleIle  209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   933  CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACCGCTGGCATC  992

Qy   210  SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspGlyGln  229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   993  TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG  1052

Qy   230  IleLeuAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu  249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATCAGGATGCGGACGATCACACCAAGCCTG  1112

Qy   250  ValAspGluGluLysLysAlaSerAsnProAspPheProGluValSerGlyIleTyrVal  269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GTGGATGAGGTGAAGGCCAGCAACCCGGACTTCCCAGAGGTGTCAGGCAGTGGAATTTATGTG  1172
```

```
Qy    270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAsnSerGlyAspIleIle 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 CAAGAGGTTGCCCCGAATTCACCTTCTCAGAGAGGCGGCCATCCAAAGATGGTGACATCATC 1232

Qy    290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 GTCAAGGTCAACGGGCGGCCGTTCCTCGAGTCCAGCGAGCTGCAGGAGGCCGTGCTGACC 1292

Qy    310 GluSerProLeuLeuLeuGluValAlaArgArgAlaAspAsnAspGlyLeuAsnProSerIleAla 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GAGTCTCCTCTCCTTACTCGAGGTGCGGGGGGGAACGACGACCTCTCCTTCAGCCATCGCA 1352

Qy    330 ProGluValValMet 334
          |||||||||||||
Db   1353 CCTGAGGTGGTCATG 1367
```

```
RESULT 8

AAA57361

ID    AAA57361 standard; cDNA; 1563 BP.
XX
AC    AAA57361;
XX
DT    03-OCT-2000   (first entry)
XX
DE    cDNA encoding a HtrA-2 (high temperature requirement A-2) protein.
XX
KW    HtrA-2; high temperature requirement A-2 protein; TANGO 219; arthritis;
KW    bone disorder; cartilage disease; bone breakage; cartilage tearing;
KW    osteoporosis; osteoarthritis; bone wearing; ischemic heart disease;
KW    atherosclerosis; hyperplasia; ventricular fibrillation; visceromegaly;
KW    placentomegaly; cardiac defect; Beckwith-Weidermann syndrome; cancer;
KW    cellular migration; cell proliferation; serine protease;
KW    Alzheimer's disease; multiple sclerosis; Huntington's disease;
KW    Parkinson's disease; neurodegeneration; frontal lone dementia;
KW    cortical lewy body disease; acquired immunodeficiency syndrome; AIDS;
KW    dementia; chronic epilepsy; adipocyte function; heart disease;
KW    ischemic heart disease; ventricular fibrillation; myocardial infarction;
XX
OS    Mus sp.
XX
FH    Key         Location/Qualifiers
FH
FT    CDS         268..1314
FT                /*tag= a
FT                /product= "HtrA-2 (high temperature requirement A-2)
FT                protein"
FT    variation   replace(396,C)
FT                /*tag= b
FT                /note= "the encoded amino acid is changed to Asp; this
FT                variation is present only when the other two variations
FT                are absent"
FT                /*tag= c
FT    variation   replace(426,C)
FT                /note= "the encoded amino acid is changed to Asp; this
FT                variation is present only when the other two variations
FT                are absent"
FT                /*tag= c
FT    variation   replace(498,C)
FT                /note= "the encoded amino acid is changed to Asp; this
FT                variation is present only when the other two variations
FT                are absent"
FT                /*tag= d
FT    variation   replace(498,C)
FT                /note= "the encoded amino acid is changed to Asp; this
FT                variation is present only when the other two variations
FT                are absent"
XX
CC    The present sequence encodes a murine HtrA-2 (high temperature
CC    requirement A-2) protein. The specification also describes TANGO 219
CC    polypeptides. HtrA-2 nucleic acid, proteins and its modulators can be
CC    used to treat bone and/or cartilage associated diseases or disorders,
CC    e.g. bone breakage, cartilage tearing, osteoporosis, arthritis,
CC    osteoarthritis, and bone wearing. HtrA-2 is also used to treat ischemic
CC    heart disease, atherosclerosis, hyperplasia, and ventricular
CC    fibrillation. HtrA-2 can also modulate IGF function and can be used to
CC    treat visceromegaly, placentomegaly, cardiac and adrenal defects,
CC    Beckwith-Weidermann syndrome, and cancer. It is useful to treat disorders
CC    involving abnormal cellular migration, proliferation. HtrA-2 acts as a
CC    serine protease and can be used to treat multiple sclerosis, HtrA-2 can
CC    also be used to treat Alzheimer's disease, multiple sclerosis,
CC    Parkinson's disease, neurodegeneration, frontal lone dementia, cortical
CC    lewy body disease, acquired immunodeficiency syndrome (AIDS), dementia
CC    and chronic epilepsy. HtrA-2 nucleic acid, proteins and its modulators
CC    can modulate adipocyte function and adipocyte-related processes and
CC    disorders. TANGO 219 can be used to treat heart diseases or conditions,
CC    e.g., ischemic heart disease or atherosclerosis, or cerebrovascular
CC    accidents, or more particularly, treating or preventing conditions
CC    involving heart contraction, and the impulse generating nodes and cardiac
CC    muscle cells, e.g. ventricular fibrillation or myocardial infarction
XX
SQ    Sequence 1563 BP; 337 A; 462 C; 458 G; 306 T; 0 U; 0 Other;
```

```
US-10-617-443B-2 (1-334) x AAA57361 (1-1563)

Alignment Scores:
Pred. No.:
Score:                      1.36e-159
                            1635.00
Percent Similarity:         100.00%
Percent Similarity:         100.00%
Best Local Similarity:      97.38%
Query Match:                3

DB:
```

Length: 1563
Matches: 325
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```
Qy     10 GlyLeuHisGlnLeuSerSerProArgThrTyrLysPheAsnPheIleAlaAspValValGlu 29
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 GGTCTCCACCAGCTCGAGCAGCCCCGGCTACAAAGTTCAACTTCATTGCTGACGTGGTGGAG 396

Qy     30 LysIleAlaProAlaValValHisIleGlLeuPheLeuArgHisProLeuPheGlyArg 49
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 AAGATCGCACCAGCCGTGGTCCACATAGGCTCTTCCCTGAGACACCCGCTGTTTGGCCGC 456

Qy     50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 AACGTGCCCCTGTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 516

Qy     70 AsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnIleGlnLeuGln 89
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCGCCCCGGGACGCCAGCTCAAGGTGCAG 576

Qy     90 LeuGlnAsnGlyAspSerTyrGluAlaThrLysGlyAspIleAspProSerSerArgIle 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    577 CTACAGAATGGGGACTCCTATGAGGCCACCAAAGGAGACATCGACAAAGAAGTTCGGACAATT 636

Qy    110 AlaThrIleLysHisProLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    637 GCCACCATCAAGCATCCCAAGAAAAAGTTCCCTGTGTTGTTGCTAGGGCACTCGGCC 696

Qy    130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
PI    Barnes TM;
XX
PN    WPI; 2000-452366/39.
DR    P-PSDB; AAY93963.
XX
DR    P-PSDB; AAY93963.
XX
PT    Novel human high temperature requirement-2 and TANGO 219 secreted
PT    proteins and nucleic acids encoding them which are useful for treating
PT    atherosclerosis, placentomegaly, cancer, Alzheimer's disease and
PT    diabetes.
XX
PS    Claim 2; Fig 8A-C; 122pp; English.
XX
```

```
PA    (MILL-) MILLENNIUM PHARM INC.
PR    98US-00224246.
PR    30-DEC-1998;
PF    29-DEC-1999;   99WO-US031157.
PD    06-JUL-2000.
PN    WO200039149-A2.
```

Db   697 GACCTCCGGCCTGGGGAGTTTGTGGTGGCCATCGGCGAGTCCTTCGCCCTACAGAACACA 756
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   150 ValThrThrGlyIleValSerThrAlaLeuHisArgGluLeuGlyLeuArgGluArgAsp 169
Db   757 GTGACAACGGGCACCATCGTCAGCACTGCCCAGCGGGAGCTGGGCCTCCGGGAC 816

Qy   170 SerAspMetAspTyrIleGlnThrArgAspAlaIleIleLeuAsnTyrGlyAsnSerGlyGlyPro 189
Db   817 TCCGACATGGACTACATCCAGACACGGGATGCCATCATCCTCAACTACGGGAACTCCGGGGGACCA 876

Qy   190 LeuValAsnLeuAspGlyValIleGlyIleLeuAsnThrLeuLysValThrAlaGlyIle 209
Db   877 CTGGTGAACCTGGATGGCGAGGTCATTGGGATCAACACGCTCAAGGTCACCGCTGGCATC 936

Qy   210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAsnGln 229
Db   937 TCCTTTGCCATCCCCTCAGACCGCATCACGCGTTCCTCACGGAGTTCCAAGACAAGCAG 996

Qy   230 IleLysAspTrpLeuValArgPheLeuGlyIleIleArgMetArgThrIleThrProSerLeu 249
Db   997 ATCAAAGACTGGCTGGTCCGGTTCCTAGGGATCATACGGATGCGGACGATCACCACCAAGCCTG 1056

Qy   250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValValSerSerGlyIleTyrVal 269
Db 1057 GTGGATGAGCTGAAGGCCAGCAACCCGGACTTTCCCGAGGTGGTCAGCAGTGGAATTTATGTG 1116

Qy   270 GlnGluLeuValAlaProAlaAsnSerProSerGlnArgGlyGlyIleGlnArgAspGlyAspIleIle 289
Db 1117 CAAGAGGTTGCGCCGGCGAATTCACCTTCTCAGAGAGGGGGCATCCAAGAGGTGGTGACATCGATC 1176

Qy   290 ValLysSerValAsnGlyValArgProLeuValAlaAspSerSerGluLeuGlnLeuAlaValLeuThr 309
Db 1177 GTCAAGGTCAACGGCGGCGTCCTCTAGTGGACTCGAGTGAGCTGCAGAGGGCCGTGCTGACC 1236

Qy   310 GluSerProLeuLeuGluLeuGluGlyValArgArgGlyTyrAsnAspAspProLeuLeuProPheSerIleAla 329
Db 1237 GAGTCTCCTCTGCTGGAGCTGGAGGGCGTGCGGCGCGGGACAGACGACCCTCTTCTCAGCATCGCA 1296

Qy   330 ProGluValValMet 334
          |||||||||||||||
Db 1297 CCTGAGGTGGTCATG 1311

RESULT 9

ACF16993

ID   ACF16993 standard; cDNA; 1583 BP.
XX
AC   ACF16993;
XX
DT   15-SEP-2003 (first entry)
XX
DE   Human NOV24a protein encoding cDNA SEQ ID NO:109.
XX
KW   Human; NOVX; G protein-coupled receptor; cytostatic; cardiovascular;
KW   immunosuppressive; anti-HIV; antiasthmatic; antiarteriosclerotic; AIDS;
KW   hypotensive; gene therapy; cardiomyopathy; atherosclerosis; hypertension;
KW   congenital heart defect; aortic stenosis; atrial septal defect; neoplasm;
KW   atrioventricular canal defect; pulmonary stenosis; prostate cancer;
KW   uterine cancer; graft versus host disease; multiple sclerosis; GPCR;
KW   acquired immunodeficiency syndrome; Crohn's disease; bronchial asthma;
KW   chromosome mapping; forensic identification; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20029568-A2.
XX
PD   14-NOV-2002.
XX
PR   02-MAY-2002; 2002WO-US014341.
XX
PR   03-MAY-2001; 2001US-0289935P.
PR   07-MAY-2001; 2001US-0290087P.
PR   08-MAY-2001; 2001US-0289620P.
PR   08-MAY-2001; 2001US-0289621P.
PR   09-MAY-2001; 2001US-0289917P.
PR   09-MAY-2001; 2001US-0289918P.
PR   11-MAY-2001; 2001US-0290194P.
PR   14-MAY-2001; 2001US-0290753P.
PR   15-MAY-2001; 2001US-0291189P.
PR   15-MAY-2001; 2001US-0291243P.
PR   16-MAY-2001; 2001US-0291243P.
PR   18-MAY-2001; 2001US-0292001P.
PR   21-MAY-2001; 2001US-0292374P.
PR   22-MAY-2001; 2001US-0292587P.
PR   23-MAY-2001; 2001US-0293107P.
PR   24-MAY-2001; 2001US-0293589P.
PR   25-MAY-2001; 2001US-0293747P.
PR   29-MAY-2001; 2001US-0294110P.
PR   30-MAY-2001; 2001US-0294434P.
PR   14-AUG-2001; 2001US-0312192P.
PR   17-AUG-2001; 2001US-0313173P.
PR   17-AUG-2001; 2001US-0313187P.
PR   12-SEP-2001; 2001US-0318728P.
PR   12-SEP-2001; 2001US-0318744P.
PR   15-NOV-2001; 2001US-0335910P.
PR   28-NOV-2001; 2001US-0335910P.
PR   28-NOV-2001; 2001US-0333891P.
PR   03-JAN-2002; 2001US-0333942P.
PR   04-JAN-2002; 2002US-0345220P.
PR   01-MAY-2002; 2002US-00136071.
XX

PI   Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Casman SJ;
PI   Ediger SR, Ellerman K, Gangolli EA, Gerlach VL, Gorman L;
PI   Gunther E, Herrmann JL, Ji W, Lepley DM, Lewin DA, Li L;
PI   Macdougall JR, Malyankar UM, Mezes PD, Padigaru M, Patturajan M;
PI   Peyman JA, Rastelli L, Rieger DK, Rothenberg ME, Shenoy SG;
PI   Smithson G, Spytek KA, Stone DJ, Taupier RJ, Tcherney VI;
PI   Vernet CAM, Voss EZ, Zerhusen BD, Zhong H, Miller CE;
XX
DR   WPI; 2003-1111987/10.
DR   P-PSDB; ABR96184.
XX
PT   New NOVX polypeptides and polynucleotides useful for treating or
PT   preventing e.g. cardiomyopathy, atherosclerosis, hypertension, congenital
PT   heart defects, aortic stenosis, atrial septal defect, or atrioventricular
PT   canal defect.
XX
PA   (CURA-) CURAGEN CORP.
XX
PS   Claim 20; Page 197; 491pp; English.
XX
CC   ACF16939 to ACF17000 encode the human G protein-coupled receptor (GPCR)
CC   proteins, designated NOVX proteins, given in ABR96130 to ABR96191. The
CC   NOVX sequences can have cytostatic, cardiovascular, antiasthmatic,
CC   immunosuppressive, anti-HIV (human immunodeficiency virus), hypotensive
CC   and antiarteriosclerotic activities, and can be used in gene therapy.
CC   NOVX polypeptides can be used for treating a syndrome associated with a
CC   human disease such as a pathology associated with the polypeptide. NOVX
CC   polypeptides, polynucleotides and antibodies can be used for treating or
CC   preventing e.g. cardiomyopathy, atherosclerosis, hypertension, congenital
CC   heart defects, aortic stenosis, atrial septal defect, atrioventricular
CC   canal defect, pulmonary stenosis, prostate cancer, uterine cancer,
CC   neoplasm, graft versus host disease, multiple sclerosis, acquired
CC   (AIDS), Crohn's disease, multiple sclerosis, or bronchial asthma. The
CC   nucleic acid sequences may be used in chromosome mapping, identifying
CC   individual from minute biological samples (tissue typing), and in
CC   forensic identification of a biological sample. ACF17001 to ACF17717
CC   represent PCR primers and probes for the NOVX sequences, which are used
CC   in an example from the present invention
XX
SQ   Sequence 1583 BP; 272 A; 556 C; 490 G; 264 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:
Pred.:   1.39e-159   Length: 1583
Score: 1635.00   Matches: 325
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 97.38%   Indels: 0

RESULT 10

```
DB:                         10                          Gaps:          0

US-10-617-443B-2 (1-334) x ACF16993 (1-1583)

Qy    10  GlyLeuHisGlnLeuSerSerProArgThrYrLysPheAsnPheIleAlaAspValValGlu   29
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   580  GGTCTCCACCAGCTGAGCAGCCCGGCGTTACAAGTTCAACTTCATTGCTGACGTGGTGGAG  639

Qy    30  LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg    49
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   640  AAGATCGCCCCAGCCGTGGTCCACATAGAGCTCTTCTTGAGACACCCGCTGTTTGGCCGC   699

Qy    50  AsnValProLeuSerGlyLySerGlyPheIleMetSerGluGlyAlaGlyLeuIleLeuThr   69
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   700  AACGGTGCCCCTGTCCAGCGGCGGTTCTGGCTTCATCACATGTCAGAGGGCGCTGGTATCACC  759

Qy    70  AsnAlaHisValValSerSerAspSerAlaAlaProGlyArgGlnLeuLeuLysValGln    89
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   760  AATGCCCACGTGGTGTCCAGCAACGACGCGCTGCCCGGCCGGCAGCTCAAGGTGCAG      819

Qy    90  LeuGlnAsnGlyAspSerArgTyrGluAlaThrIleLeuLysAspAspIleAspLysSerAspIle  109
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   820  CTACAGAATGGGGACTCCTATGAGGCCACCATAAAGACATCGACAAGAAGTCGGACATT    879

Qy   110  AlaThrIleLysTyrIleGluHisIleProLeuGlySLeuProValLeuLeuGlyHisSerAla  129
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   880  GCCACCATCAAGATCCATCCACAGAAAAAGCTCCCTGTTGTTGCTGGGTCACTCGGCC     939

Qy   130  AspLeuArgProGlyTyrGluProPheValAlaLeuIleGlyLySerProPheAlaLeuGluGlnAsnThr  149
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   940  GACCTGCGGCCCGGCTACGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA  999

Qy   150  SerPheAlaIleProSerArgAspArgIleThrArgPheLeuThrHisGluProGlnAsp   169
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1000  TCCTTTGCCATCCCCTCAGACCGGTCCTCAGAGTTCCAAGACGCAGGAGGCAG         1059

Qy   170  SerAspMetAspTyrIleGluGlnThrAspAspAlaAlaIleLeuIleLeuAsnTyrGlyAsnSerGlyGlyPro  189
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1060  TCCGACATGGACTACATCCAAGACGATGCCAGTGCCATCACTACACGGGGTCACTCGGGGGGACCA  1119

Qy   190  LeuValAsnLeuAspGlyGluValAlaIleGlyTyrIleLeuAsnThrLeuLySValThrAlaGlyYrIle  209
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1120  CTGGTGAACCTGGATGGGGAGGTTGCCATCAACACGCTCAAGGTCACGGCTGGGCAATC  1179

Qy   210  ValAspGlyGluValAlaSerAspProSerSerGlyAsnProAspLeuYrIleLeuTyrYrVal  229
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1180  GTGGATGAGGTGAGCTCAGACCCCAGAAGTGGGAATCCCGACCTTCCCAGGGGTGAATTTATTTGTG  1239

Qy   230  IleLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu    249
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1240  ATCAAAGACTGGAAGAAGCGCTTCATCGGCATACGGATGCGGACGATCACACCAAGCCTG  1299

Qy   250  ValAspGlyValLeuGlyAlaSerAsnProAspPheProGluValSerSerGlyYrIleTyrVal  269
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1300  GTGGATGGGGTCCTCGGGGCCAGCAACCCGGACTTTCCCGAGGTGTCGCAGTGGAATTTATTGTG  1359

Qy   270  GlnGluValAlaProAsnSerProSerGluGlnIleArgGlyTyrIleGlnAspGlyAspIle  289
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1360  CAAGAGGTTGCCGCCGAATTCACCTTCTCAGGAGGGGCATCAAGATGGTGACATCATC   1419

Qy   290  ValLySValAsnGlyArgProLeuValAspSerSerGluLeuGlnGlyValAlaValLeuThr  309
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1420  GTCAAGGTCAACGGGCGGCCTCTAGTGGACTCGAGCGAGCTGCAGGGAGTGGCCGTGCTGACC  1479

Qy   310  GluSerProLeuLeuLeuGluGlyValArgGlyTyrAsnAsppAspLeuLeuPheSerIleIleAla  329
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1480  GAGTCTCCTCTCCTTACTGGAGGTGCGGGGGAACGACGACCTCCTCTTCAGCATCGCA  1539

Qy   330  ProGluValValMet  334
              |||||||||||||||
Db  1540  CCTGAGGTGGTCATG  1554
```

AAD23855
```
ID   AAD23855 standard; cDNA; 1797 BP.
XX
AC   AAD23855;
XX
DT   07-MAR-2002 (first entry)
XX
DE   Human protease PRTS-14 cDNA.
XX
KW   Human; protease; PRTS-14; tranquilliser; gene therapy; vaccine; allergy;
KW   infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;
KW   atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;
KW   gastrointestinal disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;
KW   epithelial disorder; neurological disorder; Parkinson's disease; drug screening;
KW   hypertension; Cushing's syndrome; seizure; glaucoma; stroke;
KW   cardiant; cell proliferative disorder; multiple sclerosis; osteoporosis;
KW   diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;
KW   autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;
KW   developmental disorder; reproductive disorder; infertility; diarrhoea;
KW   dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT
FT   CDS             181..1542
FT                   /tag= a
FT                   /product= "Human protease PRTS-14 protein"
FT   sig_peptide     181..231
FT                   /tag= b
FT   mat_peptide     232..1539
FT                   /tag= c
FT                   /product= "Mature human protease PRTS-14 protein"
XX
PN   WO200183775-A2.
XX
PD   08-NOV-2001.
XX
PF   04-MAY-2001; 2001WO-US014651.
XX
PR   04-MAY-2000; 2000US-0202082P.
PR   11-MAY-2000; 2000US-0203566P.
PR   17-MAY-2000; 2000US-0205803P.
PR   25-MAY-2000; 2000US-0207477P.
PR   01-JUN-2000; 2000US-0209402P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Delegeane AM, Lal P, Hafalia A, Patterson C, Walia NK, Kearney L;
PI   Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;
PI   Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DAM;
PI   Reddy R, Yue H, Tang YT;
XX
DR   WPI; 2002-034518/04.
DR   P-PSDB; AAE14349.
XX
PT   Novel human proteases and polynucleotides encoding the proteases, useful
PT   for treating, diagnosing or preventing cell proliferative,
PT   cardiovascular, autoimmune/inflammatory, neurological and developmental
PT   disorders.
XX
PS   Claim 5; Page 150-151; 151pp; English.
XX
CC   The invention relates to human proteases (PRTS1-14) and its corresponding
CC   cDNA molecules. Human PRTS and its nucleic acid molecule are useful for
CC   the diagnosis, treatment and prevention of disorders associated with
CC   increased or decreased expression of PRTS. Examples of such disorders
CC   include, cell proliferative disorders (arteriosclerosis, atherosclerosis,
CC   hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders
CC   (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,
CC   diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,
CC   trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and
CC   viral, bacterial, fungal, parasitic, protozoal and helminthic infections)
CC   ; cardiovascular disorders (myocardial infarction, ischaemic heart
```

```
CC  disease and hypertension); neurological disorders (epilepsy, Alzheimer's
CC  disease, Pick's disease, Huntington's disease, dementia, Parkinson's
CC  disease, stroke, mental disorders including mood, anxiety and seasonal
CC  affective disorder and prion diseases); gastrointestinal disorders
CC  (Crohn's disease, anorexia, nausea, diarrhoea and jaundice); epithelial
CC  disorders (contact dermatitis, eczema, acne vulgaris, alopecia, scabies,
CC  insect bites and urticaria); reproductive disorder (infertility,
CC  disruption of estrous and menstrual cycle and gynaecomastia); and
CC  developmental disorders (renal tubular acidosis, Cushing's syndrome,
CC  seizure disorders, congenital glaucoma and cataract). PRTS is also in
CC  useful is gene therapy PRTS and its immunogenic fragments are useful for
CC  screening libraries of compounds in several drug screening assays. The
CC  present sequence is human protease PRTS-14 cDNA
XX
SQ  Sequence 1797 BP; 317 A; 621 C; 564 G; 295 T; 0 U; 0 Other;
```

```
US-10-617-443B-2 (1-334) x AAD23855 (1-1797)

                                               Length:       1797
                                               Matches:      325
                                               Conservative: 0
                                               Mismatches:   0
                                               Indels:       0
                                               Gaps:         0

Qy  10 GlyLeuHisGlnLeuSerSerProArgThrTyrLeuLysPheAsnPheIleAlaAspValValGlu 29
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 565 GGTCTCCACCAGCTGAGCAGCCCCGGCTACACAAGTTCAAACTTCATTGCTGACGTGGTGGAG 624

Qy  30 LysIleAlaProAlaAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 625 AAGATCGCACCAGCCGCTGGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 684

Qy  50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 685 AACCTGCCCCTGTCCAGCGGGTTCTGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC 744

Qy  70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyTyrArgGlnGlnLeuLysValGln 89
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 745 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCCGGGCCAGAGTTCATTGCTGACGTGCAG 804

Qy  90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysGlyAspLysSerSerAspIle 109
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 805 CTACAGAATGGGACTCCTATGAGGCCACCATCAAAAGACATCGACAAGAGCTCGGACATT 864

Qy 110 AlaThrIleLysIleHisProLysGlyLysSerLeuProValLeuLeuLeuGlyHisSerAla 129
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 865 GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCC 924

Qy 130 AspLeuArgProGlyGlyPheValAlaIleAlaIleGlySerProPheAlaLeuGlnAsnThr 149
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 925 GACCTGCGGCCTGGGGGAGTTTGTGGCTATCGCCATCGGGTCCCCTTCGCCCTACAGAACACA 984

Qy 150 ValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyLeuArgAsp 169
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 985 GTGACAACCGGCATCGTCAGCACTGCCCAGGGAGGGAGCTGGGCCTCCGGAC 1044

Qy 170 SerAspMetAspThrIleGlnThrAspAlaAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1045 TCCGACATGGACTACATCCAGACCGATGCCGCCATCATCAACTACGGCAACTCCGGGGGCCCA 1104

Qy 190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1105 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC 1164

Qy 210 SerPheAlaIleProSerAspAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1165 TCCTTTGCCATCCCCTCAGACGACCGGATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 1224

Qy 230 IleLysAspTrpLeuLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1225 ATCAAAGACTGGGAGAAGCGCTTCATCGGCATACGGATGCGGACCATCACACCAAGCCTG 1284
```

```
RESULT 11
AAZ52362

ID  AAZ52362 standard; cDNA; 2040 BP.
XX
AC  AAZ52362;
XX
DT  24-JUL-2000  (first entry)
XX
DE  NSEQ gene-14 associated with matrix remodelling.
XX
KW  NSEQ gene; matrix-remodeling gene; Incyte clone 2814981; cancer;
KW  matrix-remodeling disease; cardiomyopathy; arthritis; angiogenesis;
KW  diabetic necrosis; atherosclerosis; fibrosis; ulceration; cytostatic;
KW  cardioactive; antiarthritic; angiogenic; antiarteriosclerotic; antiulcer;
KW  ss.
XX
OS  Homo sapiens.
XX
FH  Key             Location/Qualifiers
FT  misc_feature    1..1023
FT                  /tag= a
FT                  /note= "Shows, 60%-70% sequence identity with human serine
FT                  protease mRNA"
FT  misc_feature    3..1043
FT                  /tag= b
FT                  /note= "Protein encoded by this region shows 61% sequence
FT                  identity with an osteoblast-like cell derived protein"
XX
PN  WO200021986-A2.
XX
PD  20-APR-2000.
XX
PF  06-OCT-1999;   99WO-US023315.
XX
PR  09-OCT-1998;   98US-0016928.9.
XX
PA  (INCY-) INCYTE PHARM INC.
XX
PI  Walker MG, Volkmuth W, Klingler TM;
XX
DR  WPI; 2000-317934/27.
XX
PT  Protein co-expressed with matrix-remodeling proteins, useful in the
PT  diagnosis and treatment of cancer, cardiomyopathy, arthritis,
PT  angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and
PT  ulceration.
PS  Claim 1; Page 46-47; 55pp; English.
PT
CC  The present sequence is NSEQ gene that is co-expressed with one or more
CC  known matrix-remodeling genes in a number of biological samples using an
CC  expression vector. This sequence was identified from the Incyte clone
```

```
Qy 250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1285 GTGGATGAGCTGAAGGCCAGCAACCCCGACTTCCCAGAGGTCAGCAGTGGAATTTACGTG 1344

Qy 270 GlnGluValAlaProAsnSerProSerGlnAsnArgGlyGlyIleGlnAsnArgGlyIleIle 289
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1345 CAAGAGGTTGCGCCGAATTCACCTTCTCAGAACAGGGGCGGGATTCAGAACCGAGGCATCATC 1404

Qy 290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1405 GTCAAGGTCAACGGGCGGCCTCTAGTGGACTCTGAGCTGCAGGAGGCCGTGCTGACC 1464

Qy 310 GluSerProLeuLeuLeuValAlaArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1465 GAGTCTCCTCTCCTACTGGAGGTGCGCGGGGGGAACGACGACCTCCTCTTCAGCATCGCA 1524

Qy 330 ProGluValValMet 334
        |||||||||||||||
Db 1525 CCTGAGGTGGTCATG 1539
```

```
CC   281498l. The gene, protein, and antibody sequences can be used in the
CC   diagnosis, and treatment or prevention of a disease associated with its
CC   altered expression. The diseases that can be treated are matrix-
CC   remodeling diseases, including cancer, cardiomyopathy, arthritis,
CC   angiogenesis, diabetic necrosis, atherosclerosis, fibrosis, and
CC   ulceration.
XX
SQ   Sequence 2040 BP; 427 A; 635 C; 602 G; 376 T; 0 U; 0 Other;
DB:

US-10-617-443B-2 (1-334) x AAZ52362 (1-2040)

Alignment Scores:
Pred. No.:
Score:                 1.96e-159        Length:       2040
Percent Similarity:    1635.00          Matches:      325
Best Local Similarity: 100.00%          Conservative: 0
Query Match:           100.00%          Mismatches:   0
DB:                    97.38%           Indels:       0
                       3                Gaps:         0

Qy   10  GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu  29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72  GGTCTCCACCAGCTGAGCAGCCCCCGGCCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 131

Qy   30  LysIleAlaProAlaValValHisIleGlyLeuPheLeuArgHisProLeuPheGlyArg   49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132  AAGATCGCACCAGCCGTGGTCCACATAGGACTCTTTCTCCGACACCCGCTGTTTGGCCGC  191

Qy   50  AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleLeuThr   69
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  192  AACGTGCCCCTGTCCAGCGGGTTCGGCTTCATCATGTCAGAGGCCGGGCTGATCATCACC  251

Qy   70  AsnAlaHisLeuValValSerSerAsnSerAlaAlaProGlyArgGlnLeuLeuLysValGln 89
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  252  AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGCCGGCAGCTCCTCAAGGTGCAG  311

Qy   90  LeuGlnAsnGlyYAspSerTyrGluIleuAlaThrIleThrIleLysSerAspIle      109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312  CTACAGAATGGGAGACTCCTATGAGGGCCACCATCAAAGACATCGACAAGAGTCGGACATT 371

Qy  110  AlaThrIleLysIleHisProLysGlyGlyLeuProValLeuLeuLeuGlyHisSerArgAla 129
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  372  GCCACCATCAAGATCCATCCAAAAAAGCTCCCTGTTGGTCACTCGGCC             431

Qy  130  AspLeuArgProGlyGluThrPheValValAlaIleLeuGlySerProPheAlaLeuGlnAsnThr 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  432  GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCCGGCTCCCTTCGGCCCTACAGAACACA 491

Qy  150  ValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyArgGlyLeuArgAsp     169
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492  GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGAGAGCTGGGCCTTCGGGAC     551

Qy  170  SerAspMetAspTyrIleGlnThrAspAlaIleIleGlyAsnTyrGlyAsnSerGlyTyrPro 189
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  552  TCCGACATGGACTACATCCAGACCGATGCCATCATCAACTACGGCAACTCCGGCGGGACCA 611

Qy  190  LeuValAsnLeuAspGlyGluValIleSerAsnThrLeuUlyUlyValThrAlaGlyIle   209
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  612  CTGGTGAACCTGGATGGCGAGGTCATTGCCAACACCCTCAAGGTCACGGCTGGCATC     671

Qy  210  SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluGlnAspLysGln     229
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  672  TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG  731

Qy  230  IleLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu   249
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  732  ATCAAAGACTGGAAGAAGCGCTTCATCGGCATACCGGATGGGACCATCACACCAAGCCTG  791

Qy  250  ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal   269
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  792  GTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTGTCACGGGTTGAATTTATGTG 851

Qy  270  GlnGluValAlaProAsnSerProSerGlnArgGlyIleGlnAspGlyAspIleIle      289
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  852  CAAGAGGTTGCCGCCGAATTCACCTTCTCAGAGAGCGCCATCCAAGATGGTGGACATCATC 911

Qy  290  ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnLeuUlaValLeuThr   309
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  912  GTCAAGGTGCAACGGGGCCGTCCCTTAGTGGACTCCAGCGAGCTGCAGTTGGCTGACC    971

Qy  310  GluSerProLeuLeuLeuGluValArgArgGlyLysAsnAspAspIleuLeuPheSerIleAla 329
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  972  GAGTCTCCTCTTACTGGAGGTGCGGCGCGGGGAAAGAACGACGACCTCCTCTTCAGCATCGCA 1031

Qy  330  ProGluValMet 334
         ||||||||||||
Db 1032  CCTGAGGTGGTCATG 1046

RESULT 12
AAL42458
ID   AAL42458 standard; DNA; 2040 BP.
XX
AC   AAL42458;
XX
DT   11-JUL-2002  (first entry)
XX
DE   Human matrix-remodeling-associated nucleotide 14.
XX
KW   Human; dm; matrix-remodeling gene; extracellular matrix; gene;
KW   matrix-remodeling-associated nucleotide; screening;
KW   matrix remodelling-associated disease; angiogenesis; arthritis;
KW   atherosclerosis; cancer; cardiomyopathy; diabetic necrosis; fibrosis;
KW   ulceration.
OS   Homo sapiens.
XX
PN   US2002019000-A1.
XX
PD   14-FEB-2002.
XX
PF   26-MAR-2001; 2001US-00818143.
PR   09-OCT-1998;  98US-00169289.
XX
PA   (WALK/) WALKER M G.
PA   (VOLK/) VOLKMUTH W.
PA   (KLIN/) KLINGLER T M.
XX
PI   Walker MG, Volkmuth W, Klingler TM;
XX
DR   WPI; 2002-338319/37.
XX
PT   New isolated polynucleotide coexpressed with matrix-remodeling genes,
PT   useful in diagnosis, prognosis, prevention and treatment of diseases
PT   associated with matrix-remodeling such as angiogenesis, arthritis and
PT   cancer.
XX
PS   Claim 2; Page 29-30; 63pp; English.
XX
CC   The invention comprises human nucleotide sequences which are co-expressed
CC   with matrix-remodeling genes. Matrix-remodeling is associated with the
CC   construction, destruction and reorganisation of extracellular matrix
CC   components. The matrix-remodeling-associated nucleotides of the invention
CC   are useful for screening for and purifying ligands that specifically bind
CC   to the nucleotides of the invention. The matrix-remodeling-associated
CC   nucleotides of the invention are also useful in the diagnosis, prognosis,
CC   prevention, treatment and evaluation of therapies for diseases associated
CC   with matrix remodelling (e.g. angiogenesis, arthritis, atherosclerosis,
CC   cancer, cardiomyopathy, diabetic necrosis, fibrosis and ulceration). The
CC   present DNA sequence represents a human matrix-remodeling-associated
CC   nucleotide of the invention
XX
SQ   Sequence 2040 BP; 427 A; 635 C; 602 G; 376 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:      1.96e-159        Length:    2040
            1635.00          Matches:   325
```

```
Percent Similarity:      100.00%          Conservative:  0
Best Local Similarity:   100.00%          Mismatches:    0
Query Match:             97.38%           Indels:        0
DB:                      6                Gaps:          0

US-10-617-443B-2 (1-334) x AAL42458 (1-2040)

Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrArgLysPheAsnPheIleAlaAspValValGlu  29
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72 GGTCTCCACCAGCTGAGCAGCCCGCGCTACAGAGTTCAACTTCATTGCTGACGTGGTGGAG 131

Qy   30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg    49
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC   191

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGlyValAlaGlyIleLeuIleThr  69
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  192 AACGTGCCCCTCTCCAGCGGGTTCTCGGCTTCATCATGTCAGAGGGCCGGCCTGATCATCACC 251

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln    89
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  252 AATGCCCACGTGGTGTCCAGCAACACTGCTGCCCCCGGCCGCAGCTCAAGGTGCAG     311

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaLeuThrIleLeuAspLysLysSerAspIle     109
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312 CTACAGAATGGGGACTCCTATGAGGCCCACCATCAAAGACATCGACAAGAAGTCGGACATT 371

Qy  110 AlaThrIleLysIleLeuHisProLysGlyLysGlyLeuProValLeuLeuLeuGlyHisSerAla 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 GCCACCATCAAGATCCTACATCCCAAGAAAAAGGCTCCCTGTGTTGTTGCTGGGTCACTCGGCC 431

Qy  130 AspLeuArgProGlyIleGluPheValAlaIleAlaIleGlyLysSerProPheAlaLeuGlnAsnThr 149
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  432 GACCTGCGGCCTGGGGAGTTTGTGGCCATCGGGAAGTCCCTTCGCCTTCACAGACACA    491

Qy  150 ValThrThrGlyIleValSerThrArgLeuGlnArgGlyGlyArgGluLeuGlyTyrLeuArgAsp 169
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492 GTGACAACGGGCATCGTCAGCACTCGCCTCCAGCGGGGCTGGGCCTTCGGGGAC       551

Qy  170 SerAspMetAspTyrIleGlnThrArgAspAlaIleAlaIleLeuGluAsnTyrGlyAsnSerGlyGlyPro 189
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  552 TCCGACATGGACTACATCCAGACGGAGGCCATCATCAACTACGGGAAACTCCGGGGGGACCA 611

Qy  190 LeuValAlaAsnLeuAsnAspGlyGluValIleLeuGlyIleLeuAsnThrLeuLysValThrAlaGlyIle 209
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  612 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC  671

Qy  210 SerPheAlaIleProSerArgAspArgIleThrArgPheLeuThrGluPheGlnAspLysGlyGln 229
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  672 TCCTTTGCCATCCCCTCACGGGACCGCATCACGCGGTTCCTCACAGAGTTCCAAGACAAGCAG 731

Qy  230 IleLeuSerAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu    249
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  732 ATCAAAAGACTGGAAGAAGCGCTTCATCGGGACTATCGGGACGATCACACCAAGCCTG     791

Qy  250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal   269
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  792 GTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTCAGCAGTGGAATTTATGTG  851

Qy  270 GlnGluValAlaProAsnSerProGlnArgGlyIleGlnAspGlyAspIleIle       289
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  852 CAAGAGGTTGCGCCGAATTCACCTTCTCAGAGAGGGCATCCAAGATGGTGACATCATC   911

Qy  290 ValLysValAlaAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAlaValLeuThr 309
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  912 GTCAAGGTCAACGGGCGGTCCTCTAGTGGACTCGAGTGAGCTGCAGGAGGCCGTGCTGACC  971

Qy  310 GluSerProLeuLeuLeuGluGluValArgGlyLysAsnAspAspLeuLeuPheSerIleAla   329
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  972 GAGTCTCCTCTACTGGAGGTGCGGGGGAAGAACGATGACCTCCTCTTCAGCATCGCA     1031

Qy  330 ProGluValValMet 334
        |||||||||||||||
Db  330
```

Db   1032 CCTGAGGTGGTCATG 1046

```
Db   1032 CCTGAGGTGGTCATG 1046

RESULT 13
ABK51497

ID   ABK51497 standard; cDNA; 2094 BP.
XX
AC   ABK51497;
XX
DT   13-AUG-2002  (first entry)
XX
DE   cDNA encoding human LP protein LP241.
XX
KW   Human; LP241; inflammatory disorder; diabetes; bone disease;
KW   cardiovascular disease; male reproductive system disease; osteoporosis;
KW   Paget's disease; myeloma; Alzheimer's disease; contraceptive;
KW   liver cancer; growth factor-mediated disease; anaphylaxis; coagulation;
KW   sepsis; skeletal muscle dystrophy; asthma; breast cancer; gene; ss.
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            206..1567
FT                  /*tag= a
FT                  /product= "Secreted protein, LP241"
XX
PN   WO200232939-A2.
XX
PD   25-APR-2002.
XX
PF   10-OCT-2001; 2001WO-US027759.
XX
PR   19-OCT-2000; 2000US-0241813P.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Lu D,  Song HY,  Su EW,  Wang H;
XX
DR   WPI; 2002-454591/48.
DR   P-PSDB; AAU97040.
XX
PT   New secreted human LP polypeptides or polynucleotides, useful for
PT   treating mammals suffering from conditions associated with aberrant
PT   levels of an LP polypeptide, e.g. cancers, osteoporosis, Paget's disease
PT   or Alzheimer's disease.
XX
PS   Claim 1; Page 144-146; 148pp; English.
XX
CC   The invention relates to isolated human polypeptides designated LP102,
CC   LP187, LP190 and LP241, and the polynucleotides encoding them. The LP
CC   polypeptide or the LP polypeptide agonist is useful for treating a mammal
CC   suffering from a disease, condition or disorder associated with aberrant
CC   levels of an LP polypeptide. The LP polypeptides or polynucleotides, or
CC   the antibodies are useful for treating, preventing or diagnosing cancers,
CC   inflammatory disorders, bone diseases or cardiovascular diseases. In
CC   particular, the LP102 polypeptides, polynucleotides or antibodies are
CC   useful for diagnosing, preventing or treating male reproductive system
CC   diseases, osteoporosis, Paget's disease, metastatic or myeloma associated
CC   bone diseases or Alzheimer's disease. These are also useful as
CC   contraceptive agents. The LP187 polypeptides, polynucleotides or
CC   antibodies are particularly useful for diagnosing, preventing or treating
CC   liver cancer, as well as other growth factor-mediated diseases and
CC   conditions. The LP190 polypeptides, polynucleotides or antibodies are
CC   useful for diagnosing, preventing or treating asthma, anaphylaxis, and
CC   diseases related to coagulation or sepsis. LP241 polypeptides,
CC   polynucleotides or antibodies are useful especially for diagnosing,
CC   preventing or treating skeletal muscle dystrophy, breast cancer or
CC   diabetes. The LP polynucleotides are also useful for constructing DNA
CC   vectors that may be employed in medicine, or for preparing the proteins.
CC   The present sequence represents the coding sequence of LP241
XX
SQ   Sequence 2094 BP; 360 A; 737 C; 633 G; 364 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:
Score: 2.03e-159
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 97.38%
DB:

Length: 2094
Matches: 325
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-617-443B-2 (1-334) x ABK51497 (1-2094)

```
Qy    10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   590 GGTCTCCACCAGCTGGAGCAGCCCGCGTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 649

Qy    30 LysIleAlaProAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   650 AAGATCGCACCAGCCGTGCTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGGCGC 709

Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyValIleIleThr 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   710 AACGTGCCCCTGTCCAGCGGGTTCATCATGTCAGAGGCCGGGGTCATCATCACC 769

Qy    70 AsnAlaHisValSerSerAsnSerAlaAlaProGlyArgGlnIleLeuLysValGln 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   770 AATGCCCACGTGTCTTCCAGCAACAGTGCTGCCCCGGGAGGCAGCTCAAGGTGCAG 829

Qy    90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspLysLysGlySerAspIle 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   830 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACAAGAAGGGGAGTGGACATT 889

Qy   110 AlaThrIleLysIleHisProLeuPheGlyLeuProValLeuLeuLeuGlyHisSerAla 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   890 GCCACCATCAAGATCCATCCCCCAAGAAAAAAGCTCCCTGTGTTGTTGGTCACTCGGCC 949

Qy   130 AspLeuArgProGlyGluPheValAlaIleAlaGlySerProPheAlaLeuGlnAsnThr 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   950 GACCTCCGCCCTGGGGAGTTTGTGGCCATCGCGGGTCCCTTGGCCCTTACAGAACACA 1009

Qy   150 ValThrThrGlyTyrIleValSerThrAlaGlnAsnArgGluLeuGlyLeuArgAsp 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1010 GTGACAACGGGGCATCGTCAGCACTGCCCAGCCGGAGGGCAGGGAGCTGGGCTGGCAT 1069

Qy   170 SerAspMetAspTyrIleGlnThrArgAspAlaIleIleLeuAsnTyrGlyAsnSerGlyGlyPro 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1070 TCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGGAACTCCGGGGGACCA 1129

Qy   190 LeuValAsnLeuAspGlyValIleGlyTyrIleAsnThrLeuLysValThrAlaGlyIle 209
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1130 CTGGTGAACCTGGATGGCGAGGTCATTGGCTACATCAACACGCTCAAGGTCACGGCTGGCATC 1189

Qy   210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1190 TCCTTTGCCATCCCCTCGGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 1249

Qy   230 IleLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1250 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATACGGATGCGGACGATCACCACCAAGCCTG 1309

Qy   250 ValAspGlyLeuGlyLysAlaAserAsnProAspPheProGluValSerGlyIleTyrVal 269
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1310 GTGGATGGAGCTGAAGGCCAGCAACCCCGACTTCCCAGAGGTCAGTGGAATTTATGTG 1369

Qy   270 GlnGluValAlaProAsnSerProSerGlnArgGlyTyrIleGlnIleAspGlyAspIleIle 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1370 CAAGAGGTTGCGCCGAATTCACCTTCACCTCAGAGATTGGTGACATCATC 1429

Qy   290 ValLysValValAsnGlyArgProLeuValAspSerGlnLeuGlnIleAlaValLeuThr 309
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1430 GTCAAGGTCAACGGCGTCCCTCTAGTGGACTCGGAGCTCCAGGGGAGGCCCGTGCTGACC 1489

Qy   310 GluSerProLeuLeuLeuGluValArgGlyAsnAspAspLeuLeuPheSerIleAla 329
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1490 GAGTCTCCTCTCCTACTGGAGGTGCGGGGGAACGACGACCTCCTCTTCAGCATCGCA 1549
```

Qy   330 ProGluValValMet 334
         ||||||||||||||
Db 1550 CCTGAGGTGGTCATG 1564

Human PRSP clone encoding sequence #1.

RESULT 14
ACF12810
ACF12810 standard; DNA; 2543 BP.
ID   ACF12810 standard; DNA; 2543 BP.
XX
Db 1550 CCTGAGGTGGTCATG 1564

Key   Location/Qualifiers
CDS   1..2543
         /*tag= a
         /product= "PRSP"

WO2003011905-A1.
13-FEB-2003.
30-JUL-2002; 2002WO-AU001010.
30-JUL-2001; 2001AU-00006707.
(PRIN-) PRINCE HENRY'S INST MEDICAL RES.

Nie G,   Salamonsen LA,   Li Y,   Hampton AL,   Findlay JK;

WPI; 2003-268108/26.
P-PSDB; ABR91169.

Claim 5; Fig 3a; 156pp; English.

The present sequence relates to a new isolated nucleic acid molecule,
which is expressed in endometrium and placenta and is upregulated in
pregnant uterus and is highly expressed during placental development,
encodes a protein having serine protease activity and has an insulin-like
growth factor (IGF)-binding motif. The compound is considered
gynecological, cytostatic and cardiant. The enzyme is specifically
expressed in association with embryo implantation and placentation in a
pregnant uterus. The nucleic acid is useful for preparing a composition
for treating PRSP-related condition e.g., infertility, endometriosis,
cancer or a disease of the heart, testis or ovaries. Further, it is
useful for detecting, diagnosing or monitoring a condition involving a
change in PRSP expression. The sequence is present in the exemplification
of the specification

New nucleic acid encoding a protein having serine protease activity and
an insulin-like growth factor-binding motif, useful for preparing a
composition for treating a pregnancy-related serum protease-related
condition e.g., infertility.

Sequence 2543 BP; 458 A; 858 C; 777 G; 450 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.65e-159
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 97.38%
DB:

Length: 2543
Matches: 325
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-617-443B-2 (1-334) x ACF12810 (1-2543)

Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29

RESULT 15

ID   AD001092 standard; DNA; 2554 BP.
XX
AC   AD001092;
XX
DT   29-JUL-2004 (first entry)
XX
DE   Human serine protease HTRA3 related DNA, SEQ ID 7.
XX
KW   Cytostatic; Serine protease inhibitor; Apoptosis stimulator; human;
KW   serine protease; enzyme; HTRA3; pancreatic cancer; cancer; ds.
XX
OS   Homo sapiens.
XX
PN   WO200439407-A1.
XX
PD   13-MAY-2004.
XX
PF   30-OCT-2003; 2003WO-JP013920.
XX
PR   01-NOV-2002; 2002JP-00320075.
PR   27-JAN-2003; 2003JP-00017892.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PI   Horikoshi K, Kitahara O, Watanabe T, Taniyama Y, Nishizawa S;
XX
PT   Apoptosis inducers inhibiting activity or expression of serine protease.
PT   HTRA3 for treatment, prevention and diagnosis of pancreatic cancer.
XX
PS   Disclosure; SEQ ID NO 7; 124pp; Japanese.
XX
CC   The present invention relates to apoptosis inducers and protease
CC   inhibitors, which contain a compound inhibiting the activity or
CC   expression of human serine protease HTRA3 (AD001090 and AD001115). The
CC   invention is useful for the treatment, prevention and diagnosis of
CC   pancreatic cancer, and also of other cancers including colon, breast,
CC   lung, prostate, oesophagus, stomach, liver, spleen, kidney, gall bladder,
CC   ovary, testis, bladder, thyroid and brain cancer and blood neoplasms. The
CC   present sequence was used to illustrate the invention.
XX
DR   WPI; 2004-400073/37.
XX
SQ   Sequence 2554 BP; 447 A; 866 C; 786 G; 455 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 1635.00   Length: 2554
Percent Similarity: 100.00%   Matches: 325
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 97.38%   Mismatches: 0
DB: 12   Indels: 0
                                           Gaps: 0

US-10-617-443B-2 (1-334) x AD001092 (1-2554)

Qy   10 GlyIleLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleLeuAlaAspValGlu 29
         GGTCTCCACCAGCTGAGCAGCCCGCTACAAGTTCAACTTCATTGGACTGGGTGGAG
Db   606 GGTCTCCACCAGCTGAGCAGCCCGCTACAAGTTCAACTTCATTGGACTGGGTGGAG 665

Qy   30 LysIleAlaProAlaValValHisIleGluLeuArgHisProLeuPheGlyArg 49
         AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCGACACCCGCTGTTTGGCCGC
Db   666 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCGACACCCGCTGTTTGGCCGC 725

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGlyAlaGlyLeuIleLeuThr 69
         AACGTGCCCCTGTCCAGCGGGTTCTGCTTCATCATGTCAGGAGCGGGCCTGATCATCACC
Db   726 AACGTGCCCCTGTCCAGCGGGTTCTGCTTCATCATGTCAGGAGCGGGCCTGATCATCACC 785

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnInLeuLysValGln 89
         GCCAACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTCGGGGTCACTCGGCC
Db   786 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCGCCCCGGGACGGCAGCTCAAGGTGCAG 845

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLeuLysAspIleLysSerSerAspIle 109
         CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCAAAAGCAGCGAC ATT
Db   846 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCAAAAGCAGCGACATT 905

Qy   110 AlaThrIleLeuSerIleHisProLysGlyLysGlyLeuProValLeuLeuLeuGlyHisSerAla 129
         GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTCGGGGTCACTCGGCC
Db   906 GCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTCGGGGTCACTCGGCC 965

```
Qy  130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  966 GACCTGCGGCCTTGGGGAGAGTTTGTGGTGGCCATCGGGCAGTCCCTTCGCCTACAGAACACA 1025

Qy  150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1026 GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGCAGGGAGCTGGGCCTCCGGGAC 1085

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaIleIleIleAsnTyrGlyAsnSerGlyGlyPro 189
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1086 TCCGACATGGACTACATCCAGACGGATGCCATCATCATCAACTACGGGGAAACTCCGGGGGGACCA 1145

Qy  190 LeuValAsnLeuAspGlyGluValIleIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1146 CTGGTGAACCTGGATGGGGAGGTCATTGGCATCAACACCCTCAAGGTCACGGCCTGGCATC 1205

Qy  210 SerPheAlaIleProSerArgAspArgIleThrArgPheLeuThrGluGlnAspLysGln 229
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1206 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAAGACAAGCAG 1265

Qy  230 IleLeuLysAspTrpLysArgPheIleGlyIleLeuGlyMetArgThrIleThrProSerLeu 249
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1266 ATCAAAGACTGGAAGAGGCCTTCATCGGCCATACGGACGATCACCACCAAGCCTG 1325

Qy  250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleLeuTyrVal 269
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1326 GTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTCAGCAGTGGAATTTATGTG 1385

Qy  270 GlnGluValAlaProAlaSerGlnIleAsnArgGlyGlyIleIleGlnAsnAspGlyIleIle 289
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1386 CAAGAGGTTGCGCCCGGAATTCACCTTCTCAGAGAGGGGCATCCAAGATGGTGACATCATC 1445

Qy  290 ValLysValAlaSerGlyArgProLeuValAspSerGluGlnLeuAlaValLeuThr 309
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1446 GTCAAGGTCACGGGCGGCTCCCTTAGTGGACTCGGAGTGGGCTGACAGGACGCCGTTGACC 1505

Qy  310 GluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1506 GAGTCTCCTCCTTACTGAGGTGCGCGGCGGGAACGACGACCTCCTTCAGCATCGCA 1565

Qy  330 ProGluValValMet 334
            |||||||||||||||
Db 1566 CCTGAGGTGGTCATG 1580
```

Search completed: June 29, 2005, 21:40:04
Job time : 956.732 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 21:11:10 ; Search time 297.971 Seconds
(without alignments)
1834.127 Million cell updates/sec

Title: US-10-617-443B-2
Perfect score: 1679
Sequence: 1 MHLALPASAGLHQLSSPRYK...........LEVRGNDDLLFSIAPEVVM 334

Searched: 1202784 seqs, 818138359 residues

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10617443/runat_29062005_135932_26096/app_query.fasta_1.718
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-DB=Issued_Patents_NA -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10617443 @CGN_1_1_93 @runat_29062005_135932_26096 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
-FGAPOP=10 -YGAPEXT=0.5 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	% Match	ID	Description
1	1034.5	61.6	1521	4		US-09-949-016-4643	Sequence 4643, Ap
2	1034.5	61.6	2036	3		US-09-949-016-399	Sequence 399, App
3	1034.5	61.6	2205	3		US-08-888-077A-41	Sequence 41, Appl
4	1027.5	61.2	2036	3		US-08-923-454A-17	Sequence 17, Appl
5	920	54.8	1859	3		US-09-724-864-22	Sequence 22, Appl
6	882	52.5	1835	3		US-08-923-454A-7	Sequence 7, Appli
7	882	52.5	2040	4		US-09-075-460-4	Sequence 4, Appli
8	882	52.5	2187	3		US-08-923-454A-23	Sequence 23, Appl
9	882	52.5	2187	3		US-09-923-454A-24	Sequence 24, Appl
10	882	52.5	2476	3		US-09-008-271A-23	Sequence 23, Appl
11	878	52.3	2187	3		US-08-923-454A-30	Sequence 30, Appl
12	834.5	49.7	1787	3		US-08-923-454A-3	Sequence 3, Appli
13	834.5	49.7	2551	3		US-08-923-454A-26	Sequence 26, Appl
14	779	46.4	1503	3		US-08-923-454A-5	Sequence 5, Appl
15	779	46.4	2144	3		US-08-923-454A-28	Sequence 28, Appl
16	579.5	34.5	539	1		US-08-322-742-18	Sequence 18, Appl
17	454	27.0	1230230	4		US-09-438-185A-1	Sequence 1, Appl
18	450.5	26.8	1425	4		US-09-489-039A-6523	Sequence 6523, Ap
19	444	26.4	1332	4		US-09-902-540-4672	Sequence 4672, Ap
20	444	26.3	24754	4		US-09-902-540-1230	Sequence 1230, Ap
21	441.5	26.3	1368	4		US-09-711-164-245	Sequence 245, Appl
22	437	26.0	1386	4		US-09-902-540-4598	Sequence 4598, Ap
23	437	26.0	27707	4		US-09-902-540-1226	Sequence 1226, Ap
24	431	25.7	1068	4		US-09-711-164-246	Sequence 246, App
25	430	25.6	1230025	4		US-09-438-185A-1	Sequence 1, Appl1
26	428	25.5	1436	3		US-09-199-637A-131	Sequence 131, Appl1
27	425	25.3	1428	4		US-09-252-991A-12923	Sequence 12923, A
28	425	25.3	1455	4		US-09-252-991A-12561	Sequence 12561, A
29	424	25.3	732	3		US-08-923-454A-1	Sequence 1, Appli
30	422	25.1	1980	1		US-08-350-741-1	Sequence 1, Appli
31	422	25.1	1980	2		US-08-463-875A-1	Sequence 1, Appli
32	418	24.9	1500	4		US-09-673-898-5	Sequence 5, Appli
33	417	24.8	1083	4		US-09-489-039A-6463	Sequence 6463, Ap
34	415	24.7	1110	4		US-09-673-898-7	Sequence 7, Appli
35	414	24.7	1242	4		US-09-388-090-5	Sequence 5, Appl
36	414	24.7	1326	4		US-09-388-089B-13	Sequence 13, Appl
37	414	24.7	1395	4		US-09-388-089B-10	Sequence 10, Appl
38	414	24.7	1395	4		US-09-388-090-3	Sequence 3, Appl
39	414	24.7	1611	1		US-08-485-569-1	Sequence 1, Appli
40	414	24.7	1611	1		US-08-480-993-1	Sequence 1, Appli
41	414	24.7	1611	2		US-07-903-079B-1	Sequence 1, Appli
42	413	24.6	1830121	4		US-09-557-884-1	Sequence 1, Appli
43	413	24.6	1830121	4		US-09-643-990A-1	Sequence 1, Appli
44	411.5	24.5	1488	4		US-09-489-039A-5255	Sequence 5255, Ap
45	411.5	24.5	1494	4		US-09-489-039A-5193	Sequence 5193, Ap

ALIGNMENTS

RESULT 1
US-09-949-016-4643

; Sequence 4643, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 4643
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4643

Qy

Alignment Scores:
Pred. No.:
Score: 1034.50
Percent Similarity: 80.43%
Best Local Similarity: 62.08%
Query Match: 61.61%
DB:

Length: 1521
Matches: 203
Conservative: 60
Mismatches: 55
Indels: 9
Gaps: 2

US-10-617-443B-2 (1-334) x US-09-949-016-4643 (1-1521)

Qy 10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29

```
               |||            :::|||   |||:::|||:::||||||||||||||||||||||||
Db           2 GGGCAGGAAGATCCCAACAGTTTGCGCCATAAATATAACTTTATCGCGGACGTGGTGGAG 61

Qy          30 LysIleAlaProAlaValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
               ||||||||||||||||||||||||||||||||||||   :::   |||         |||
Db          62 AAGATCGCCCCTGCCGTGGTTCATATCGAATTGTTTCGCAAGCTTCCGTTTTCTAAACGA 121

Qy          50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
                  ||||||::::::||||||||||||||||||:::||||||   |||||||||:::|||
Db         122 GAGGTGCCGGTGGCTAGTGGGTCTGGGTTTATTGTGTCGGAAGATGGACTGATCGTGACA 181

Qy          70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
               |||||||||||||||::::::                  :::   ::::::||||||:::
Db         182 AATGCCCACGTGGTGACCAAC------------------AAGCACCGGGTCAAAGTTGAG 223

Qy          90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysLysSerAspIle 109
               |||:::||||||   :::|||||||||   |||||||||:::|||:::|||:::||||||
Db         224 CTGAAGAACGGTGCCACTTACGAAGCCAAAATCAAGGATGTGGATGAGAAAGCAGACATC 283

Qy         110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
               |||   |||||||||      :::   ||||||||||||||||||||||||   |||:::
Db         284 GCACTCATCAAAATTGACCACCAGGGCAAGCTGCCTGTCCTGCTGCTTGGCCGCTCCTCA 343

Qy         130 AspLeuArgProGlyGluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
               :::||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db         344 GAGCTGCGGCCGGGAGAGTTCGTGGTCGCCATCGGAAGCCCGTTTTCCCTTCAAAACACA 403

Qy         150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGlyLeuArgAsp 169
               ||||||||||||||||||||||||   ||||||   |||:::|||||||||||||||:::
Db         404 GTCACCACCGGGATCGTGAGCACCACCCAGCGAGGCGGCAAAGAGCTGGGGCTCCGCAAC 463

Qy         170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         464 TCAGACATGGACTACATCCAGACCGACGCCATCATCAACTATGGAAACTCGGGAGGCCCG 523

Qy         190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         524 TTAGTAAACCTGGACGGTGAAGTGATTGGAATTAACACTTTGAAAGTGACAGCTGGAATC 583

Qy         210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
               |||||||||||||||||||||:::|||   :::||||||||||||      |||:::|||
Db         584 TCCTTTGCAATCCCATCTGATAAGATTAAAAAGTTCCTCACGGAGTCCCATGACCGACAG 643

Qy         230 IleLysAsp---------TrpLysLysArgPheIleGlyIleArgMetArgThrIleThr 246
                  |||               ||||||::::::|||||||||||||||   ::::::|||
Db         644 GCCAAAGGAAAAGCCATCACCAAGAAGAAGTATATTGGTATCCGAATGATGTCACTCACG 703

Qy         247 ProSerLeuValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGly 266
                  |||         |||||||||      :::   |||||||||:::|||   ||||||
Db         704 TCCAGCAAAGCCAAAGAGCTGAAGGACCGGCACCGGGACTTCCCAGACGTGATCTCAGGA 763

Qy         267 IleTyrValGlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGly 286
                  |||:::   ||||||   |||::::::|||::::::   |||||||:::::::::
Db         764 GCGTATATAATTGAAGTAATTCCTGATACCCCAGCAGAAGCTGGTGGTCTCAAGGAAAAC 823

Qy         287 AspIleIleValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnGluAla 306
               |||:::|||:::   :::||||||:::   :::|||   ::::::::::::   :::
Db         824 GACGTCATAATCAGCATCAATGGACAGTCCGTGGTCTCCGCCAATGATGTCAGCGACGTC 883

Qy         307 ValLeuThrGluSerProLeuLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPhe 326
               :::      ||||||   |||   :::   |||||||||||||||:::|||::::::
Db         884 ATTAAAAGGGAAAGCACCCTGAACATGGTGGTCCGCAGGGGTAATGAAGATATCATGATC 943

Qy         327 SerIleAlaProGluValVal 333
               ::::::   ||||||   :::
Db         944 ACAGTGATTCCCGAAGAAATT 964

RESULT 2
US-09-949-016-399
; Sequence 399, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 399
;   LENGTH: 2036
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-399

Alignment Scores:
Pred. No.:              2.57e-120      Length:         2036
Score:                  1034.50        Matches:        206
Percent Similarity:     80.18%         Conservative:   61
Best Local Similarity:  61.86%         Mismatches:     57
Query Match:            61.61%         Indels:         10
DB:                     4              Gaps:           2

US-10-617-443B-2 (1-334) x US-09-949-016-399 (1-2036)

Qy           4 AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPhe 23
               |||   ||||||:::   |||            :::|||   |||:::|||:::||||||
Db         503 GCGGAGCCTGCGGCCAAGGGCAG-GAAGATCCCAACAGTTTGCGCCATAAATATAACTTT 561

Qy          24 IleAlaAspValValGluLysIleAlaProAlaValValHisIleGluLeuPheLeuArg 43
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||   :::
Db         562 ATCGCGGACGTGGTGGAGAAGATCGCCCCTGCCGTGGTTCATATCGAATTGTTTCGCAAG 621

Qy          44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63
                  |||         |||   ||||||::::::||||||||||||||||||:::||||||
Db         622 CTTCCGTTTTCTAAACGAGAGGTGCCGGTGGCTAGTGGGTCTGGGTTTATTGTGTCGGAA 681

Qy          64 AlaGlyLeuIleIleThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArg 83
                  |||||||||:::||||||||||||||||||::::::                  :::
Db         682 GATGGACTGATCGTGACAAATGCCCACGTGGTGACCAAC------------------AAG 723

Qy          84 GlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIle 103
                  ::::::||||||:::|||:::||||||   :::|||||||||   |||||||||:::
Db         724 CACCGGGTCAAAGTTGAGCTGAAGAACGGTGCCACTTACGAAGCCAAAATCAAGGATGTG 783

Qy         104 AspLysLysSerAspIleAlaThrIleLysIleHisProLysLysLysLeuProValLeu 123
               |||:::|||:::|||||||||   |||||||||      :::   |||||||||||||||
Db         784 GATGAGAAAGCAGACATCGCACTCATCAAAATTGACCACCAGGGCAAGCTGCCTGTCCTG 843

Qy         124 LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValValAlaIleGlySerPro 143
               |||||||||   |||::::::|||||||||||||||||||||||||||||||||||||||
Db         844 CTGCTTGGCCGCTCCTCAGAGCTGCGGCCGGGAGAGTTCGTGGTCGCCATCGGAAGCCCG 903

Qy         144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArg 163
               |||:::||||||||||||||||||||||||||||||||||||   ||||||   |||:::
Db         904 TTTTCCCTTCAAAACACAGTCACCACCGGGATCGTGAGCACCACCCAGCGAGGCGGCAAA 963

Qy         164 GluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyr 183
               |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db         964 GAGCTGGGGCTCCGCAACTCAGACATGGACTACATCCAGACCGACGCCATCATCAACTAT 1023

Qy         184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1024 GGAAACTCGGGAGGCCCGTTAGTAAACCTGGACGGTGAAGTGATTGGAATTAACACTTTG 1083

Qy         204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223
               |||||||||||||||||||||||||||||||||||||||:::|||   :::|||||||||
Db        1084 AAAGTGACAGCTGGAATCTCCTTTGCAATCCCATCTGATAAGATTAAAAAGTTCCTCACG 1143
```

RESULT 3

US-08-888-077A-41
Sequence 41, Application US/08880077A
Patent No. 6020143
GENERAL INFORMATION:
  APPLICANT: ST. GEORGE-HYSLOP, PETER H
  APPLICANT: ROMMENS, JOHANNA M
  APPLICANT: FRASER, PAUL E
  TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
  TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND PROTEINS RELATED.
  NUMBER OF SEQUENCES: 41
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
  STREET: 600 SOUTH AVENUE WEST
  CITY: WESTFIELD
  STATE: NJ
  COUNTRY: USA
  ZIP: 07090-1497
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: ASCII
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/888,077A
  FILING DATE: 03-JUL-1997
  CLASSIFICATION: 530
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/592,541
  FILING DATE: 26-JAN-1996
  ATTORNEY/AGENT INFORMATION:
  NAME: PALISI, THOMAS M
  REGISTRATION NUMBER: 36,629
  REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP CIP IV
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (908) 654-5000
  TELEFAX: (908) 654-7866
  INFORMATION FOR SEQ ID NO: 41:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 2205 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 1..2205
  OTHER INFORMATION: /note= "mutTM1-TM2"
  US-08-888-077A-41

```
Qy   224 GluPheGlnAspLysGlyIleIleLysAsp----------TrpLysGlyArgPheIleGlyIle 240
             |||:||         |||||||||:||:::|||||
Db  1144 GAGTCCCATGACCGACCAGGCCCAAAGGAAAGCCATCACCAGACAAGGAGAAGTATATTGGTATC 1203

Qy   241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAspSerAsnProAspPhe 260
             :||||||:|||||
Db  1204 CGAATGAGATGTCACTCACGTCCAGCAAAGACCAAAGAGCTGAAGGACCACCGGGACTTC 1263

Qy   261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280
Db  1264 CCAGACGTGATCTCAGGAGCCTATATAATTGAAGTAAATTCCTGATACCCAGCAGAAGCT 1323

Qy   281 GlyGlyIleGlnAspGlyAspIleIleIleValLysAsnGlyArgProLeuValAspSer 300
Db  1324 GGTGGTCTCAAGGACAAAGACGGTCATATAATCAGCATCATCAATGGACAGTCCGTGGTC 1383

Qy   301 SerGluLeuGlnGluValAlaValLeuThrGluSerProLeuLeuLeuGluValArgArgGly 320
Db  1384 AATGATGTCAGCGACGGTCATTAAAAGGGAAGACACCCTGAACATGGTGGTCCGCAGGGGT 1443

Qy   321 AsnAspAspLeuPheSerIleAlaProGluValVal 333
Db  1444 AATGAAGATATCATGATCACAGTGATTCCCGAAGAAATT 1482
```

Alignment Scores:
  Pred. No.:   2.93e-120
  Score: 1034.50   Length: 2205
  Percent Similarity: 80.18%   Matches: 206
  Best Local Similarity: 61.86%   Conservative: 61
  Query Match: 3   Mismatches: 57
                                         Indels: 10
                                         Gaps: 2

US-10-617-443B-2 (1-334) x US-08-888-077A-41 (1-2205)

```
Qy     4 AlaLeuProAlaAserAlaGlyValLeuHisGlnLeuSerSerProArgTyrLysPheAsnPhe 23
Db   799 -------------------------------------------------------------AAG 799

Db   758 GATGGACTGATCGGTGACAAAGCCCACGTGGTTGACCAAC
Qy    64 AlaGlyIleLeuIleLeuIleSerValSerSerAlaSerAlaAlaProGlyArg 83
Db   698 CTTCGGTTTCTAAACGAGAGGTGCCGGTGGCTAGTGGGTCTGGGTTTATTGTGTCAGGGT 757
Qy    44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlySerPheIleMetSerGlu 63
Db   638 ATCGGCCGACGTGGTGGAGAAGATCGCCCCTCGCCCTGGTTCATATCGAATTGTTTCGGAAG 697
Qy    24 IleAlaAspValAlaGlyValLeuIleLysLeuHisProLysGlyLysLysLeuProValLeu 43
Db   579 GCCGGACCCTGCGGCCGCCAAGGCAG-GAAGATCCCAACAGTTTGCGCATAAAATAACTTT 637
Qy   104 AspLysValGlySerAspSerMetAspAspTyrIleGlnThrAspAsnAlaIleLeuAsnTyr 123
Db   919 GATGAGAAAGCAGACAATCGCCACTCACTCAAAATTGACCACCAGGGAGCTGCGTGTCTG 918
Qy   124 LeuLeuGlyHisSerAlaAspLeuGluArgProGlyGluProValValAlaIleLeuGlyLysSerPro 143
Db   860 CTGCTTGGCCGCCTCTCCAGACCTGATAAGAATTCCGGACGGTGAAGGTGACTGATGGAAGCCG 919
Qy   144 PheAlaLeuGlnAsnThrValThrThrGlyIleLeuValLeuSerThrAlaGlnGlyArg 163
Db   980 TTTTCCCTTCAAAACACAGTACACCACGGGATCGTGAGCACCAGCCCAGGAAGCGGGCAAA 1039
Qy   164 GluLeuGlyLeuArgAspSerMetAspTyrIleGlnThrAspAlaIleLeuIleLeuAsnTyr 183
Db  1040 GAGCTGGGCCTCCGCAACTCAGACATGGACTACATCCAGACCGACGCCATCATCAACTAT 1099
Qy   184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleIleGlyIleIleAsnThrLeu 203
Db  1100 GGAAACTCGGGAGGACCCCGGTTAGTAAACCTGGACGGTGAAGTGATTGGAATTAACACTTTG 1159
Qy   204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223
Db  1160 AAAGTGACAGCTGGAATCTCCTTTGCAATCCCATCTGATAAGAATTCCTCACG 1219
Qy   224 GluPheGlnAspLysGlyIleIleLysAsp----------TrpLysGlyArgPheIleGlyIle 240
Db  1220 GAGTCCCATGACCGACCAGGCCCAAAGGAAAGCCATCACCAGACAAGGAGAAGTATATTGGTATC 1279
Qy   241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAspSerAsnProAspPhe 260
Db  1280 CGAATGAGATGTCACTCACGTCCAGCAAAGACCAAAGAGCTGAAGGACCACCGGGACTTC 1339
Qy   261 ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg 280
Db  1340 CCAGACGTGATCTCAGGAGCCTATATAATTGAAGTAAATTCCTGATACCCAGCAGAAGCT 1399
Qy   281 GlyGlyIleGlnAspGlyAspIleIleIleValLysAsnGlyArgProLeuValAspSer 300
Db  1400 GGTGGTCTCAAGGACAAAGACGGTCATATAATCAGCATCATCAATGGACAGTCCGTGGTC 1459
Qy   301 SerGluLeuGlnGluValAlaValLeuThrGluSerProLeuLeuLeuGluValArgArgGly 320
Db  1460 AATGATGTCAGCGACGGTCATTAAAAGGGAAGACACCCTGAACATGGTGGTCCGCAGGGGT 1519
```

```
Qy    321 AsnAspAspLeuLeuPheSerIleAlaProGluValVal 333
          |||||||||::||||||||::|||||||||  .::
Db   1520 AATGAAGAATATCATGATCACAGTGGATTCCCGAAGAAATT 1558
```

```
RESULT 4
US-08-923-454A-17
; Sequence 17, Application US/0892345A4A
; Patent No. 6004794
; GENERAL INFORMATION:
;  APPLICANT: Creasy, Caretha
;  APPLICANT: Livi, George
;  APPLICANT: Karran, Eric
;  APPLICANT: Clinkenbeard, Helen
;  APPLICANT: Browne, Michael
;  APPLICANT: Southan, Christopher
;  TITLE OF INVENTION: HUMAN SERINE PROTEASE
;  NUMBER OF SEQUENCES: 40
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: SmithKline Beecham Corporation
;  STREET: 709 Swedeland Road
;  CITY: King of Prussia
;  STATE: PA
;  COUNTRY: USA
;  ZIP: 19406
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Diskette
;  COMPUTER: IBM Compatible
;  OPERATING SYSTEM: DOS
;  SOFTWARE: FastSEQ Version 1.5
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/923,454A
;  FILING DATE:
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 60/025436
;  FILING DATE: 06-SEPT-1996
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Baumeister, Kirk
;  REGISTRATION NUMBER: 33,833
;  REFERENCE/DOCKET NUMBER: P50547
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: 610-270-5096
;  TELEFAX: 610-270-5090
;  TELEX:
;INFORMATION FOR SEQ ID NO: 17:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 2036 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: cDNA
;  HYPOTHETICAL: NO
;  ANTI-SENSE: NO
;  FRAGMENT TYPE:
;  ORIGINAL SOURCE:
;  ORIGINAL SOURCE: Feature polymorphism at 1325
```

```
Alignment Scores:
Pred. No.:                          Length:        2036
Score:                 1.98e-119    Matches:        205
Percent Similarity:    1027.50      Conservative:    61
Best Local Similarity: 79.88%       Mismatches:      58
Query Match:           61.56%       Indels:          10
DB:                    61.20%       Gaps:             2
                       3
```

```
US-10-617-443B-2 (1-334) x US-08-923-454A-17 (1-2036)

Qy    4 AlaLeuProLeuAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPhe 23
         |||    :::|:   |||||||| ::  ::  :::::||| ||:   |::::|| || |
Db  503 GCCGAGCCTGCCGCCCAGGGCAG-GAAGATCCAACAGTTTGCGCCATAAAAATAACTTT 561
```

```
Qy    24 IleAlaAspValValGluGlyValIleAlaProAlaLeuValValHisIleLeuGluPheLeuArg 43
Db   562 ATCGCGGACGTGGGGGAGGGAGTGATCGCCCCTGCCCTGGTTCATACGAGATTGTTTCGCAAG 621

Qy    44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63
Db   622 CTTCCGCTGTTTCTAAAACGAGAGGTGCCGCTGAGCAGCGGTTCTGGGTTTATTGGTCGGAA 681

Qy    64 AlaGlyLeuIleIleThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArg 83
Db   682 GATGGACTGATCGTGACAAATGCCCACGTGGTCAGCAGCAAC---------------AAG 723

Qy    84 GlnGlnLeuLysValGlnLeuLeuGlnAsnGlyAspSerTyrGlyAlaThrIleLysLysAspIle 103
Db   724 CACCGGGGTCAAAGTTGAGCTGAGAACCGGTGCCACTTACGAACGCAAATCAAGGATGTG 783

Qy   104 AspLysLysSerAspProIleAlaThrIleIleHisProLysLysLysLeuProValLeu 123
Db   784 GATGAGAAAAGCAGACAATCGCCACTCATCAAAATTGACCACCCAAGCTGCCTGTCCTG 843

Qy   124 LeuLeuGlyHisSerAlaAspLeuArgProGlyPheValValAlaIleGlyLysSerPro 143
Db   844 CTGCTTGGCCGCGTCCTCAGAGCTGCGGCCCGGAGAGTTCGGTCGCCATCGGAAGCCCG 903

Qy   144 PheAlaLeuGlnAsnThrValThrThrGlyIleValGluSerThrAlaGlnArgGluGlyTyr 163
Db   904 TTTTCCCTTCAAAACACACAGTCACCACCGGAGTGTGAGCAGACCGCCCAGCGGGAA 963

Qy   164 GluLeuGlyIleLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAlaAsnTyr 183
Db   964 GAGCTGGGGATCCTCCGCGACTCAGACATGGACTACATCCAGACGGACGCCATCATCGCAAACTAT 1023

Qy   184 GlyAsnSerGlyGlyProLeuValAlaAsnLeuAspGlyGluValIleGlyIleAsnThrLeu 203
Db  1024 GGAAACTCGGGAGGACCCCTTGTTGCAAACCTTGATGGTGAAGTGATTGGAATTAACACTTTG 1083

Qy   204 LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223
Db  1084 AAAGTGACAGCTGGAATCTCCTTTGCAATCCCATCTGATAGAATTACAAGGTTCCTCACG 1143

Qy   224 GluPheGlnAspSerGlnIleIleLysSerAsp------TrpLysArgPheIleGlyIle 240
Db  1144 GAGTTCCATGACCGACAGCGCCAAGGCCAAAGGAAAAGCCATCACCAAGAGAAGAGTATATTGGTATC 1203

Qy   241 ArgMetArgThrIleThrProSerLeuValAlaSerAlaSerAsnProAspPhe 260
Db  1204 CGAATGAGATGTCACTCACGTTCCAGCAAAGCCAAAGAGCTGAAGGACCGGCACCCGGGACTTC 1263

Qy   261 ProGluValSerSerGlyIleTyrValGlnGluValAlaAlaProSerGlnArg 280
Db  1264 CCAGACGTGATCTCAGGACGGTATATAATTGAAGTAATTCCTGATACCCCAGCAGAAGCT 1323

Qy   281 GlyGlyIleGlnAspGlyTyrAspProIleIleValLysValAsnGlyArgProLeuValAspSer 300
Db  1324 GKTGGGTTCTCAAGGAAAAACGAGCGTCATATAATTCAGCATCAATGGACAGTCCGTGGGTCTCCGCC 1383

Qy   301 SerGlyLeuGlnGluAlaValLeuThrGluSerProLeuGluLeuGluGlnValValArgArgGly 320
Db  1384 AATGAATGTCAGCGACCAGCGTCATTAAAAGGGAAAAAGCACCCTGAACATGGTGGTCCGCAGGGGT 1443

Qy   321 AsnAspAspLeuLeuPheSerIleAlaProGluValVal 333
Db  1444 AATGAAGAATATCATGATCACAGTGGATTCCCGAAGAAATT 1482
```

```
RESULT 5
US-09-724-864-22
; Sequence 22, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D
;  APPLICANT: Murison, James G.
;  TITLE OF INVENTION: Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION: by the polynucleotides and methods for their use.
```

```
FILE REFERENCE: 11000.1050UI
CURRENT APPLICATION NUMBER: US/09/724,864
CURRENT FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 22
LENGTH: 1859
TYPE: DNA
ORGANISM: Mouse

US-09-724-864-22

Alignment Scores:
Pred. No.:

Score:               920.00       Matches:         191
Percent Similarity:  67.01%       Conservative:     67
Best Local Similarity: 49.61%     Mismatches:       66
Query Match:         54.79%       Indels:           61
DB:                  3            Gaps:              4

US-10-617-443B-2 (1-334) x US-09-724-864-22 (1-1859)
```

```
Qy      4 AlaLeuProAla-----------------------SerAlaGlyLeuHisGlnLeu 14
          |||||||||||                        :::||||||
Db      1 GCGCTGCCGGCG-------------------------SerAlaGlyLeuHisGlnLeu 14
```

```
Qy     15 SerSerProArgTyrLysPheAsnPheIleAlaAspValValGlyLysIleAlaProAla 34
          |||||||||||||:::||||||||||||||||::::::::||||||||::||||||||
Db     63 TCGTCCCCGCGTTACAAATTCAATTCATTGCCGACGTGGTGGGGAAGATTGCCCCGGCA 122
```

```
Db    183 GTGGTCCACTTGCACTGTTCCGCAGATGACAGAGACACAGGTTCCCACACTGATCTAAA 242
```

```
Qy     35 ValValHisIleGluLeuPheLeuArg------ 43
          |||||::::::||| |||
Db     43 ------------------SerAlaGlyLeuHisGlnLeu 43
```

```
Qy    303 GCTTCAGGCAGCTCACAAAGCACATGGAAGTGGATCACCCTCTCACCAACCAGGAAATTCC 362
```

```
Qy     52 oLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyValLeuIleIleThrAsnAlaHi 72
```

```
Db    363 TTCCTCCAGGGGCTCTCGGGTTCATAGTGTCTGAGGATGGCTCAGCCATCATTGTCACCAATGCCCA 422
```

```
Qy     72 sValValSerSerAsnSerAlaAlaProGlyGlnGlnLeuLysValGlnLeuLeuGlnAs 92
```

```
Db    423 CGTCCTCACCAAC--------------CAGCAGAAGATCCAGGTAGAGCTCCAGAG 464
```

```
Qy    112 eLysIleHisProLysGlyLysLeuProValLeuLeuLeuGlyHisSerAlaAspLeuAr 132
Db    525 TAAGATCAGGCCAGATACTGAGCTTCCAGTACTGCTGTGGGCGGATCCTCTGACCTCCG 584
```

```
Qy    132 gProGlyTyrGluAlaThrIleLysGlySerProPheAlaLeuGlnAsnThrValThrTh 152
Db    585 GGCTGGAGAGTTTGTGGTAGCTTTGGGCAGCCCATTTTCTCTGCAGAACAACCGTGACTGC 644
```

```
Qy    152 rGlyIleValSerThrAlaGlnArgGluGlyLeuGlyArgAspSerAspMe 172
Db    645 AGGGATTGTCAGCACCACACAGAGGGGCTGGGACTGAAGATTCAGACAT 704
```

```
Qy    172 tAspTyrIleGlnThrAspAlaIleIleLysAsnTyrGlyAsnSerGlyGlyProLeuValAs 192
Db    705 AGACTATATCCAGACAGATGCCATCATTAACCATGGAAATTCTGGGGGTCCGCTGGTGGA 764
```

```
Qy    192 nLeuAspGlyValIleGlyIleAsnThrLeuLysValThrAlaGlyIleSerPheAl 212
```

```
Db    765 CTTGGATGGCGCGAGTGATTGGTATAAACACTCTGAAGGTGACTGCGAGGGATCTCTTTGC 824
```

```
Qy    212 aIleProSerArgArgIleThrArgPheLeuThrGluPheGlnAspLysGlnIleLys-- 231
Db    825 GATCCCCTCTGATCGGATCACAAGATTCCTGACAGAGTTCCAGGACAAGCAGATCAAGG 884
```

```
Qy    232 ------AspTrpLysSerArgPheIleGlyIleArgMetArgThrIleThrProSerIe 249
Db    885 CAAGGCCCCTTTGCCAGAAGAAAATACCTGGGTCTTTCGAATGCTGCCCTCTCACTCTGAACCT 944
```

```
Qy    249 uValAspGluLysLysValAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVa 269
Db    945 CCTTCAGCAAATGAAGAAGCCAAGAATCCAGAGTTCCCTGAGGTGAGTTCTGGGATTTTGT 1004
```

```
Qy    269 lGlnGlyValAlaLeuProAsnSerProSerGlnArgGlyLysIleGlnIleAspGlyAspIl 289
Db   1005 ATATGAAGTGATTCAGGGAGTCGGCTCGCCGCAAGCTCGGGGTTGAGAGACCATGATGTAAT 1064
```

```
Qy    289 eValLysValAsnGlyArgProLeuValAspSerSerGlyLeuGlyLeuGlnGluAlaValLeuTh 309
Db   1065 TGTCAGCATAAACGGGCGAAACCTGTCACCACACAACTGATGTCATTGAAGCTGTTAAGGA 1124
```

```
Qy    309 rGluSerProLeuLeuGluValArgArgGlyAsnAspAspAspLeuLeuPheSerIleAl 329
Db   1125 CAATGACTTCTCTCCATCATTGTTGCTTCGGAGGAGTCAAACCTTGTTTGTTCTGACAGTCAC 1184
```

```
Qy    329 aProGluValVal 333
             |||||:::::
Db   1185 ACCTGAAATATC 1197
```

```
RESULT 6
US-08-923-454A-7
Sequence 7, Application US/0892345454A
Patent No. 6004794
GENERAL INFORMATION:
APPLICANT: Creasy, Caretha
APPLICANT: Livi, George
APPLICANT: Karran, Eric
APPLICANT: Clinkenbeard, Helen
APPLICANT: Browne, Michael
APPLICANT: Southan, Christopher
TITLE OF INVENTION: HUMAN SERINE PROTEASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSES: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,454A
FILING DATE:
PRIOR APPLICATION DATA:
CLASSIFICATION: 435
APPLICATION NUMBER: 60/025436
FILING DATE: 06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50547
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1835 base pairs
```

```
;  TYPE:                nucleic acid
;  STRANDEDNESS:        single
;  TOPOLOGY:            linear
;  MOLECULE TYPE:       cDNA
;  HYPOTHETICAL:        NO
;  ANTI-SENSE:          NO
;  FRAGMENT TYPE:
;  ORIGINAL SOURCE:
;  FEATURE:
;    NAME/KEY:          Coding Sequence
;    LOCATION:          251..1624
;    OTHER INFORMATION:
;  US-08-923-454A-7

Alignment Scores:
Pred. No.:                4.43e-101
Score:                    882.00
Percent Similarity:       74.84%
Best Local Similarity:    52.48%
Query Match:              52.53%
DB:                       3

                    Length:        1835
                    Matches:       169
                    Conservative:  72
                    Mismatches:    71
                    Indels:        10
                    Gaps:          2

US-10-617-443B-2 (1-334) x US-08-923-454A-7 (1-1835)

Qy    15  SerSerProArgThrTyrLysPheAsnPheIleAlaAspValValGlnLysIleAlaProAla    34
          ::||||||||||:::||||||:::||
Db   671  GCTTCTCCCCGGAGTCAGTACAACTTCATCGACAGATGTGGTGGAAGAGACAGACCTGCC     730

Qy    35  ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer      54
          ::||||||||:::||||||:::||:::||
Db   731  GTGGTCTATATCGAGATCCTGACAGCACCCTTTCTTGGGCGCGAGGTCCCTATCTCG        790

Qy    55  SerGlySerGlyPheIleMetSerGluAlaGlyTyrLeuIleIleLeuThrArgAsnAlaHisIleValVal  74
          ::|||:::||||||
Db   791  AACGGCTCAGGATTCGTTGGTGGCGTGCCCAGAGGTCTTCATTGTCACCAGCGCCATGGTGTG  850

Qy    75  SerSerAsnSerAlaAlaProGlyArgGlnIleGlnLeuLeuGlnAsnGlyAsp            94
          ::|||
Db   851  GCTGAT------------CGGCGCGAGTCCGTGTGAGACTCCTAAGCGCGGAC            892

Qy    95  SerTyrGluAlaThrIleLysAspIleAlaThrIleLysIle                       114
          :::||
Db   893  ACGTATGAGGCCCTGGTCACAGCTGTGGATCCGGAGCATCGCAACGCTGAGGATT          952

Qy   135  GluPheValAlaIleAlaIleGlyGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIle 154
          ||||||||
Db  1013  GAGTTTGTTGTTGGTTGGCCATGGGAAGTCCCTTTGCACTGCAGAACACGGATCACATCCGGCATT  1072

Qy   155  ValSerThrAlaGlnAsnArgGluGlyIleArgAspSerAspMetAspTyr             174
          :::||||||:::||
Db  1073  GTTAGCTCTGCTCAGCCGTTCCAGCGACACCGGACTCCCCAAACCAATGTGGAATAC      1132

Qy   175  IleGlnThrAspAlaIleIleIleAsnTyrGlyAsnSerGlyValGlyProLeuValAsnLeuAsp 194
          ||||||||
Db  1133  ATTCAAACTGATGCCAGCTATTGATTTGGAAACTCTGGAGGTGCCGCGTTAACCTGGAT      1192

Qy   195  GlyGlyValIleGlyIleIleAsnThrLeuAlaGlyTyrIleSerPheAlaIlePro       214
          :::||
Db  1193  GGGGAGGTGATTGGAGTGAACACCATGAAGGTCAGCTGGAATCTCCTTTGCCATCCCT      1252

Qy   215  SerAspArgAlaIleThrArgPheLeuThrGluPheGlnAsnAspLysGln            230
          :::
Db  1253  TCTGATCGTGCTTCGAGAGTTTCTGCACACGTGGGGAAAAGAGAATTCCTCCCGGAATC---Ile  1312

Qy   231  LysAspTrpLysValArgPheLysThrIleThrProSerLeuVal                   250
          :::
Db  1313  AGTGGGGGTCCCAGCGGGGCGGCTCACGGGTGACCTGAGTCCCAGACATCCTT            1372

Qy   251  AspGluLeuLysAlaSerAsnProAspPheProGluValSerSerSerGlyIleTyrValGln 270
```

```
Db  1373  GCTGAACTACAGCTTCGAGAACCAAGCTTTCCGGATGTTGTGTACTCATCCAT           1432
Qy   271  GluValAlaIleAlaProAsnSerProSerGlnArgGlyValGlyIleGlnAspLeuIleIleVal  290
Db  1433  AAAGTCATCCTGGGCCTCCCCTGCACACCCGGTCGTCGGGCCCTGGTTGATGTTTG         1492
Qy   291  LysValAsnGlyArgProLeuValAspSerSerGlyLeuGlnIleAlaValLeuThrGlu    310
Db  1493  GCCCATTGGGGAGCCAGGAAATGCTGAAGAGTGTTTATGAAGGCTGTTTCGGAACCAA       1552
Qy   311  SerProLeuLeuLeuGlnValArgGlyAsnAspAspLeuPheSerIleAlaPro          330
Db  1553  TCCCAGTTGGGCCAGTGCAGATCCGGCGGGGACGAGAAACACTGACCTTATATGTGACCCCT   1612
Qy   331  GluVal 332
Db  1613  GAGGTC 1618

RESULT 7
US-09-075-460-4
; Sequence 4, Application US/09075460A
; Patent No. 6489136
; GENERAL INFORMATION:
;   APPLICANT: Zervos, Antonis S.
;   TITLE OF INVENTION: CELL PROLIFERATION RELATED GENES
;   FILE REFERENCE: 10284/004001
;   CURRENT APPLICATION NUMBER: US/09/075,460A
;   CURRENT FILING DATE: 1998-05-08
;   EARLIER APPLICATION NUMBER: US 60/046,077
;   EARLIER FILING DATE: 1997-05-09
;   NUMBER OF SEQ ID NOS: 16
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 2040
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (248)...(1834)
; US-09-075-460-4

Alignment Scores:
Pred. No.:               5.28e-101
Score:                   882.00
Percent Similarity:      74.84%
Best Local Similarity:   52.48%
Query Match:             52.53%
DB:                      4

                    Length:        2040
                    Matches:       169
                    Conservative:  72
                    Mismatches:    71
                    Indels:        10
                    Gaps:          2

US-10-617-443B-2 (1-334) x US-09-075-460-4 (1-2040)

Qy    15  SerSerProArgThrTyrLysPheAsnPheIleAlaAspValValGlnLysIleAlaProAla    34
          ::||||||||||:::||||||:::||
Db   881  GCTTCTCCCCGGAGTCAGTACAACTTCATCGACAGATGTGGTGGAAGAGACAGACCACCTGCC    940

Qy    35  ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer      54
          ::||||||||:::||||||:::||:::||
Db   941  GTGGTCTATATCGAGATCCTGACAGCGGCCACCCTTTCTTGGGCGCGAGGTCCCTATCTCG     1000

Qy    55  SerGlySerGlyPheIleMetSerGluAlaGlyTyrLeuIleIleLeuThrArgAsnAlaHisIleValVal  74
          ::|||:::||||||
Db  1001  AACGGCTCAGGATTCGTTGGTGGCGTGCCCAGAGGTCTTCATTGTCACCAACGCCCATGGTGTG  1060

Qy    75  SerSerAsnSerAlaAlaProGlyArgGlnIleGlnLeuLeuGlnAsnGlyAsp            94
          ::|||
Db  1061  GCTGAT------------CGGCGCGAGTCCGTGTGAGACTGCTAAGCGCGGAC            1102

Qy    95  SerTyrGluAlaThrIleLysAspIleAlaThrIleLysIle                       114
          :::||
Db  1103  ACGTATGAGGCCCTGGTCACAGCTGTGGATCCGGAGCATCGCAACGCTGAGGATT          1162

Qy   115  HisProLysLysLeuProValLeuLeuLeuGlnAsnGlyHisSerAlaAspLeuArgProGly 134
```

```
APPLICATION NUMBER:  US/08/923,454A
FILING DATE:         435
CLASSIFICATION:      435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  60/025436
FILING DATE:         06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME:                Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50547
TELECOMMUNICATION INFORMATION:
TELEPHONE:           610-270-5096
TELEFAX:             610-270-5090
TELEX:

INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH:              2187 base pairs
TYPE:                nucleic acid
STRANDEDNESS:        single
TOPOLOGY:            linear
MOLECULE TYPE:       cDNA
HYPOTHETICAL:        NO
ANTI-SENSE:          NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY:            Coding Sequence
LOCATION:            603...1976
OTHER INFORMATION:
US-08-923-454A-23
```

```
Alignment Scores:
Pred. No.:
Score:                     5.92e-101    Length:        2187
Percent Similarity:        882.00       Matches:       169
Best Local Similarity:     74.84%       Conservative:  72
Query Match:               52.48%       Mismatches:    71
DB:                        52.53%       Indels:        10
                                        Gaps:          2
```

```
US-10-617-443b-2 (1-334) x US-08-923-454A-23 (1-2187)
```

GENERAL INFORMATION:
Patent No. 6004794
RESULT 8
US-08-923-454A-23
; Sequence 23, Application US/08923454A

APPLICANT: Creasy, Caretha
APPLICANT: Livi, George
APPLICANT: Karran, Eric
APPLICANT: Clinkenbeard, Helen
APPLICANT: Browne, Michael
APPLICANT: Southan, Christopher
TITLE OF INVENTION: HUMAN SERINE PROTEASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:

```
Qy    175  IleGlnThrAspAlaIleIleAsnThrYrGlyAsnSerGlyGlyProLeuValAsnLeuAsp 194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1485  ATTCAAACTGATGCAGCTATTGATTTTGGAAACTCTGGAGGTCCCCTGGTTAACCTGGAT 1544

Qy    195  GlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIleSerPheAlaIlePro 214
           :|||||||||||:|||:||||:||:|:|||||||||||||||||||||||:|||
Db   1545  GGGGAGGTGATTGGAGTGAACACCATGAAAGGTCACAGCTGGAATCTCCTTTGCCATCCCT 1604

Qy    215  SerAspArgIleThrArgPheLeuThrGluPheGlnAspGln---------Ile 230
           ||||||||:||||||||||||||||:|||:|||
Db   1605  TCTGATCGTCTTCAGAGTTCTGCATCGTGGGGAAAAGAAATTCCTCCTCCGGAATC----Ile 1664

Qy    231  LysAspTrpLysSerArgPheIleGlyIleIleArgMetArgThrIleThrProSerLeuVal 250
           :|||||||||||||||||||||||||||||||||||
Db   1665  AGTGGGGTCCCAGCGCGCGCTACATTGGGGTTGATGATGCTGACCCCTGAGTCCCAGCATCCTT 1724

Qy    251  AspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrValGln 270
           |||:::|||:|||||:::|||||||
Db   1725  GCTGAAACTACAGCTTCGAGAAACCAAGCTTTCCCATGTTCAGCATGGTGTACTCATCCAT 1784

Qy    271  GluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAspGlyAspSerIleVal 290
           ||||||:||||:|||||
Db   1785  AAAGTCATCCTGGGGCCTCCCCTGCACACCGGCTGGTCTCCGGCCTGGGTGATGGTAGTTTG 1844

Qy    291  LysValAsnGlyArgProLeuValAspSerSerGlnLeuGlnLeuAlaValLeuThrGlu 310
           :|||:|||||||||:::||||||||:|||:||
Db   1845  GCCAATTGGGGAGCAGAGATGGTTACAAAATGCTGAAGATGTTTATGAAGCTGTTCGAACCCAA 1904

Qy    311  SerProLeuLeuLeuGluValAlaArgGlyAsnAspAspLeuLeuPheSerIleAlaPro 330
           :::||:|||||:|||:||||||:|||||||:::
Db   1905  TCCCAGTTGCAGTGCAGATCCGCGGGGACGAGAAAACACTGACCTTATATGTGACCCCT 1964

Qy    331  GluVal 332
           ||||||
Db   1965  GAGGTC 1970
```

```
RESULT 9
US-08-923-454A-24
; Sequence 24, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
;  APPLICANT: Creasy, Caretha
;  APPLICANT: Livi, George
;  APPLICANT: Karran, Eric
;  APPLICANT: Clinkenbeard, Helen
;  APPLICANT: Browne, Michael
;  APPLICANT: Southan, Christopher
;  TITLE OF INVENTION: HUMAN SERINE PROTEASE
;  NUMBER OF SEQUENCES: 40
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: SmithKline Beecham Corporation
;  STREET: 709 Swedeland Road
;  CITY: King of Prussia
;  STATE: PA
;  COUNTRY: USA
;  ZIP: 19406
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Diskette
;  COMPUTER: IBM Compatible
;  OPERATING SYSTEM: DOS
;  SOFTWARE: FastSEQ Version 1.5
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/923,454A
;  FILING DATE:
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 60/025436
;  FILING DATE: 06-SEPT-1996
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Baumeister, Kirk
;  REGISTRATION NUMBER: 33,833
;  REFERENCE/DOCKET NUMBER: P50547
;  TELECOMMUNICATION INFORMATION:
```

```
;  TELEPHONE: 610-270-5096
;  TELEFAX: 610-270-5090
;  TELEX:
; INFORMATION FOR SEQ ID NO: 24:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 2187 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: cDNA
;  HYPOTHETICAL: NO
;  ANTI-SENSE: NO
;  FRAGMENT TYPE:
;  ORIGINAL SOURCE:
;  FEATURE:
;  NAME/KEY: Coding Sequence
;  LOCATION: 603...1976
;  OTHER INFORMATION:
US-08-923-454A-24
```

```
Alignment Scores:
Pred. No.:
Score:                        5.92e-101
Percent Similarity:           882.00
Best Local Similarity:        74.84%
Query Match:                  52.48%
DB:                           52.53%
                              3
```

```
                              Length:           2187
                              Matches:          169
                              Conservative:     72
                              Mismatches:        71
                              Indels:            10
                              Gaps:               2
```

```
US-10-617-443B-2 (1-334) x US-08-923-454A-24 (1-2187)
```

```
Qy     15  SerSerProArgThyrLysPheAsnPheIleAlaAspValValAlaGlulysIleAlaProAla 34
           :::||||||||||:::||||||:|||||||||:|||||||||||||||:::
Db   1023  GCTTCTCCCCGGAGTCAGTACAACTTCATCGCAGATGTGGTGGCAGAACAGCACCTGCC 1082

Qy     35  ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer 54
           |||||:::|||||||:||||||||||||||||||||||:::
Db   1083  GTGGTCTTATATCGAGATCCCTGGACCGGCCACCCTTTCTTCTTCGGGCCGCCGAGGTCCCTATCTCG 1142

Qy     55  SerGlySerGlyPheIleMetSerGlyAlaGlyLeuValGlnLeuGlnAsnGlyAsp 74
           |||||||||||||||||||||||||||||||:::||
Db   1143  AACGGGCTCAGGATTCGTGGGCTGCCGGATGGGCTCATTGTCACCAACGCCCATGTGGTG 1202

Qy     75  SerSerAsnSerAlaAlaProGlyAlaArgGlnInLeuLysGlnAlaAsnGlyAsp 94
           |||:::||||||:|||||||||||||||||||||||:::
Db   1203  GCTGAT------------CGGCGCAGAGTCCGTGTGAGACTGCTAAGCGGCGCGCA 1244

Qy     95  SerTyrGluAlaThrIleLysAspIleAspIleAlaThrIleLysIle 114
           :::|||||||||:||:::
Db   1245  ACGTATGAGGCGCGGTCACAGCTGTGGGATCCCGTGGCCAGACATCCCAAACCCTGAGGATT 1304

Qy    115  HisProLysGlyLeuProValLeuLeuLeuGlyHisSerAlaAlaAspLeuArgProGly 134
           |||||||||||:::|||||||||||||||||:|||||
Db   1305  CAGACTAAGGAGGCCTCTCCCCACGGTCCCCTGGCGGCTCACAGCTGAATGTCCGGCAAGGG 1364

Qy    135  GluPheValValValAlaIleGlyLysSerProPheAlaLeuGlnAsnThrValThrGlyIle 154
           :||:|||||||||||||||:::||:::
Db   1365  GAGTTTGTTGTGTTGGCCATGGGAAGTCCCTTTGCACTGCAGAACACAGTCACATCCGGGCATT 1424

Qy    155  ValSerThrArgLeuAlaGlnArgGlyLeuArgGluArgAsnSerAspMetAspTyr 174
           |||||:::||||:::||||||||||
Db   1425  GTTAGCTCGTCGCTCAGCCTCCAGAGACCTGGGACTCCCCCAAACCAAGTGGAATAC 1484

Qy    175  IleGlnThrAspAlaIleIleAsnThrYrGlyAsnSerGlyGlyProLeuValAsnLeuAsp 194
           |||||||||||||||||||||||||||||||||||||||||||||
Db   1485  ATTCAAACTGATGCAGCTATTGATTTTGGAAACTCTGGAGGTCCCCTGGTTAACCTGGAT 1544

Qy    195  GlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIleSerPheAlaIlePro 214
           :||||||||||:||:||||:||:|:|||||||||||||||||||||||:|||
Db   1545  GGGGAGGTGATTGGAGTGAACACCATGAAAGGTCACAGCTGGAATCTCCTTTGCCATCCCT 1604

Qy    215  SerAspArgIleThrArgPheLeuThrGluPheGlnAspGln---------Ile 230
           |||||||:||||||||||||||||:|||:||
Db   1605  TCTGATCGTCTTCAGAGTTCTGCATCGTGGGGAAAAGAAATTCCTCCTCCGGAATC----Ile 1664
```

```
Qy   231  LysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeuVal 250
               :::|||||||||:::|||
Db  1665  AGTGGGTCCCAGCGGCGGCTACATTGGGGTGATGAATGCTGACCCTGAGTCCCAGCACCTT 1724

Qy   251  AspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleThrTyrValGln 270
               :::|||
Db  1725  GCTGAACTACAGCTTCGAGAACACCAAGCTTTCCGATGTTCAGCATGGTGTACTCATCCAT 1784

Qy   271  GlyValAlaProAsnSerProSerGlnArgGlyGlyGlyIleGlnAspGlyAspIleIleVal 290
               :::|||
Db  1785  AAAGTCATCCTGGGGTCCCCTGCACACCGGGCTGGTCTGCGGCCCTGGATGATGTTTG 1844

Qy   291  LysValAsnGlyAlaArgProLeuValAspSerSerGluLeuGlnLeuGlnLeuAlaValLeuThrGlu 310
               |||:::|||
Db  1845  GCCATTGGGGAGCAGAGATGGGTACAAAATGCTGAAGAGATGTTTATGAAGCTGTTGAACCCAA 1904

Qy   311  SerProLeuLeuGluLeuGlyValAlaArgGlyGlyAlaAspAspAspLeuLeuPheSerIleAlaPro 330
               |||:::||||||
Db  1905  TCCCAGTTGGCAGTGCAGATCCGGGGGGAGGGGACGAGAAACACTGACCTTATATGTGACCCCT 1964

Qy   331  GlyVal 332
               |||
Db  1965  GAGGTC 1970
```

RESULT 10
US-09-008-271A-23
Sequence 23, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT:
  Bandman, Olga
  Hillman, Jennifer L.
  Yue, Henry
  Guegler, Karl J.
  Corley, Neil C.
  Tang, Tom Y.
  Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Incyte Pharmaceuticals, Inc.
  STREET: 3174 Porter Dr.
  CITY: Palo Alto
  STATE: CA
  COUNTRY: USA
  ZIP: 94304
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: DOS
  SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/008,271A
  FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: <unknown>
  FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: Mohan-Peterson, Sheela
  REGISTRATION NUMBER: 41,201
  REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 650-855-0555
  TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
  LENGTH: 2476 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
IMMEDIATE SOURCE:
  LIBRARY: SINIUCT01
  CLONE: 2680548

SEQUENCE DESCRIPTION: SEQ ID NO: 23 :
US-09-008-271A-23

Alignment Scores:

Pred. No.:		
Score:	7.28e-101	Length: 2476
Score:	882.00	Matches: 169
Percent Similarity:	74.84%	Conservative: 72
Best Local Similarity:	52.48%	Mismatches: 71
Query Match:	52.53%	Indels: 10
DB:	3	Gaps: 2

US-10-617-443B-2 (1-334) x US-09-008-271A-23 (1-2476)

```
Qy    15  SerSerProArgTyrLysPheAsnPheIleAlaAspValValGluLysIleAlaProAla 34
               :::                          CGGCGCAGAGTCCGTGGTGGAGACTGCTAAGGCGGCGAC
Db   661  GCTGAT------------------------ 702

Qy    75  SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuGlnLeuAsnGlyAsp 94
               :::-------------------CGGCGCAGAGTCCGTGGTGGAGACTGCTAAGGCGGCGAC
Db   601  AACGGGCTCAGGATTCCGTGGGTGCCGGATGGCCTCATTGTCACCAACGCCACCGTGGTG 660

Qy    35  ValValHisIleGluLeuProLeuArgHisProGlyArgAsnValProLeuSer 54
               :::|||||:::||
Db   481  GCTTCTCCCCCGAGTCAGTACAAACTTCATCCGAGATGGGAGGAAGACACGACCTGCC 540

Qy    55  SerGlySerGlyPheIleMetSerGluValAlaGlyLeuLeuIleLeuThrAsnAlaHisValVal 74
               :::|||||:::||||||||
Db   541  GTGGTCTTATATCGAGATCCTGGACCGCGGCACCCTTTCTTGGGCGTCGGGGAGGTTCCCTATCG 600

Qy    95  SerTyrGluAlaThrIleLysLysSerAspIleAlaIleLeuLysIle 114
               :::|||||||||||:::
Db   703  ACGTATGAGGCCGTGGTCACAGCTGTGGATCCGGGACGCATCCGAAACGGCTAAGGCGATT 762

Qy   115  HisProLysLysLysGlyLeuProValLeuLeuLeuGlyHisSerAlaAlaAspLeuArgProGly 134
               :::||||||:::||||||
Db   763  CAGACTAAGGAGCCCTCTCCCCACGCTGCCTCGCCGAGACCGATCACATCCGGCCGCAAGGG 822

Qy   135  GluPheValValAlaIleIleGlySerProPheAlaLeuGluLeuGlyAsnGlyIleIle 154
               :::||||||||||
Db   823  GAGTTGTTGTTGTTGCCATTGGGAAGTCCCTTGCACTGCAGAAGACACGATTCACATCCGGCCAT 882

Qy   155  ValSerThrAlaGlnArgGlyLeuGlyLeuGluArgGlyGlyLeuGlyIleSerPheAlaIlePro 174
               :::||||||
Db   883  GTTAGCTCTGCTCAGGTCCCAGCCAGGCAGGACCTGGGAGACTCCCCCAAACCAAATGTGGAATAC 942

Qy   175  IleGlnThrAspAlaIleIleIleGluAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAsp 194
               :::|||||
Db   943  ATTCAAACTGATGCCACTATTATTGATTTTGGAAACTCTGGAGGTCCCCTGGTTAACCTGGAT 1002

Qy   195  GlyGluValIleIleGlyIleAsnThrLeuLysGlyValThrAlaGlyLysIleSerProAlaIlePro 214
               :::||||||||||
Db  1003  GGGGAGGTGATTGGGAGTGAACACCATGAAGGTCACAGCTGGAAATCTCCTTGCCATCCCT 1062

Qy   215  SerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln------------Ile 230
               :::                                             Ile
Db  1063  TCTGATCGTCTTCGAGAGTTTCTGACACGTGGGGAAAAAGAGAAATTCCTCTCCGGAATC 1122

Qy   231  LysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeuVal 250
               :::|||||||||:::|||
Db  1123  AGTGGGTCCCAGCGGCGGCTACATTGGGGTGATGAATGCTGACCCTGAGTCCCAGCACCTT 1182

Qy   251  AspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleThrTyrValGln 270
               :::|||
Db  1183  GCTGAACTACAGCTTCGAGAACACCAAGCTTTCCGATGTTCAGCATGGTGTACTCATCCAT 1242

Qy   271  GlyValAlaProAsnSerProSerGlnArgGlyGlyGlyIleGlnAspGlyAspIleIleVal 290
               :::|||
Db  1243  AAAGTCATCCTGGGGTCCCCTGCACACCGGGCTGGTCTGCGGCCCTGGATGATGTTTG 1302

Qy   291  LysValAsnGlyAlaArgProLeuValAspSerSerGluLeuGlnLeuGlnLeuAlaValLeuThrGlu 310
               |||:::|||
Db  1303  GCCATTGGGGAGCAGAGATGGGTACAAAATGCTGAAGAGATGTTTATGAAGCTGTTGAACCCAA 1362
```

RESULT 11

US-08-923-454A-30

; Sequence 30, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Karran, Eric
; APPLICANT: Clinkenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2187 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE: Polymorphic variants at 672 and 1435
; FEATURE: aa24=Arg/Cys aa278=Ala/Val
; NAME/KEY: Coding Sequence
; LOCATION: 603...1976
; OTHER INFORMATION:
US-08-923-454A-30

US-10-617-443B-2 (1-334) x US-08-923-454A-30 (1-2187)

```
Qy   311 SerProLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAlaPro 330
             |||||||||||||||||::|
Db  1363 TCCCAGTTGGCAGTGCAGATCCGCGGGGGACGAGAGAAACACTGACCTTATATGTGACCCCT 1422

Qy   331 GluVal 332
             |||||
Db  1423 GAGGTC 1428
```

US-10-617-443B-2 (1-334) x US-08-923-454A-30 (1-2187)

```
Qy    15 SerSerProArgGlyTyrLysPheAsnPheIleAlaAspValValGlnLysIleAlaProAla  34
             :::|||||:||||||::|
Db  1143 AACGGCCTCAGGATTCGTGGTGCCTCAGTACAACTTCATCGCAGATGTGGTGCAGAAGATCGCCGCC 1082

Qy    35 ValValHisIleIleGluLeuProLeuPheGlyArgHisProLeuPheGlyArgAsnAsnGlyAsp  54
             :::||||||::|:||||:::|
Db  1083 GTGGTCTCATATATCGGAGATCCTGGACCCGGACACCCTTTCTTGGGCCGAGGTCCCTATCCG 1142

Qy    75 SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuGlnLeuGlnAsnGlyAsp  94
             :::|||||||:::|
Db  1203 GCTGAT------------CCGGCGAGAGTCCGTGTGAGACTGCTAAGCCCGAC 1244

Qy    95 SerTyrGluAlaThrIleLysPheIleAspGlyLysLysSerAspIleAlaThrIleLysIle  114
             :::|||||||||::|:::|
Db  1245 ACGTATGAGCCCGTGACCAGCTGTGGGGATCCCGTGGCAGACATCGCCAACGGCTGAGGATT 1304

Qy   115 HisProLysLysLeuProValLeuLeuLeuGlyHisSerAlaAspLeuArgProGly  134
             ||||:|::||||:::||
Db  1305 CAGACTAAGGAGCCCTTCCCCACGCTGCCTCCTGGGAGACGCTCAGCCTGATGGTCCGGCAAGGG 1364

Qy   135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrAlaValThrThrGlyIle  154
             |:||::|||||||::|:|
Db  1365 GAGTTTGTTGTTGCCATCGGAAGTCCCTTTGCACTCCAGAACACAGCAGTCACAACCGGCATT 1424

Qy   155 ValSerThrAlaGlnAsnArgGlyLyGlyGluLeuGlyLeuArgAspSerAspMetAspTyr  174
             ||||:|||||::|:|
Db  1425 GTTAGCTGTTCTGCCCAGAGGGCTCCAGCCGAGAACCTGCCCCCCAAACCAATGTGGAATAC 1484

Qy   175 IleGlnThrArgAspAlaIleIleGluAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAsp  194
             :::|||||:||||||:::|
Db  1485 ATTCAAACTAGATCCGACGCTATTGATTTTGGAAAACTCTGGAGTCCCCCTGGTTAAACCTGGAT 1544

Qy   195 GlyGluValIleGlyIleAsnThrLeuLysValThrThrAlaGlyIleIleSerPheAlaIlePro  214
             ||||:||||||::|:|
Db  1545 GGGGGAGGTGATTGGAGTGAACACCATGAAGGTCCCGGAATCTCCTTTGCCATCCCT 1604

Qy   215 SerAspArgIleThrArgPheLeuThrGluPheGlnIleAspGlyGln-----------Ile  230
             :::||||||||:::|
Db  1605 TCTGATCGTATCTTCCGAGAGTTTCTGCATCGTGGGGAAAGAGAATTCCTCCTCCGGAATC  Ile 1664

Qy   231 LysAspTrpLysSerArgPheIleGlyIleAlaArgMetArgThrIleThrProSerLeuVal  250
             :::|||||:|:|||::|
Db  1665 AGTGGGTCCCAGCGCGCGCTACATTGGGGTGATGAGGATGCTGGACGTCCAGCCCAGCATCTT 1724

Qy   251 AspGluLeuLysAlaSerAsnProAspProGluProGluProValSerSerGlyIleTyrValGln  270
             :::|||||:|:|
Db  1725 GCTGAACTACAGCTTCGAGAACCACAAGCTTTCCCGATGTTCAGCATGGTGTACTCATCCAT 1784

Qy   271 GluValAlaProAlaAsnSerGlnArgGlyGlyIleGlnAsnAspProAspIleAspIleVal  290
             :::|||||:||:|
Db  1785 AAAGTCATCCTTGGGCGTCCCCTGCCACCGCCGCTGGTCGTGATGGATGATTTG 1844

Qy   291 LysValAsnGlyArgProLeuVAlAspSerSerGluGlnIleGlnLysAlaValLeuThrGlu  310
             ||||:||||||:::|
Db  1845 GCCATTGGGGAGCAGGATGGTACAAAATGCTGAAGATGTTTATGAGGCTGTTGGAACCCAA 1904

Qy   311 SerProLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAlaPro  330
             ||||||||||||:::|
Db  1905 TCCCAGTTGGCAGTGCAGATCCGCGGGGGACGAGAGAAACACTGACCTTATATGTGACCCCT 1964

Qy   331 GluVal 332
             |||||
Db  1965 GAGGTC 1970
```

RESULT 12
US-08-923-454A-3

Alignment Scores:
Pred. No.:
Score: 878.00
Percent Similarity: 74.53%
Best Local Similarity: 52.17%
Query Match: 52.29%
DB:

1.9e-100

Length: 2187
Matches: 168
Conservative: 72
Mismatches: 72
Indels: 10
Gaps: 2

```
; Sequence 3, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
;   APPLICANT: Creasy, Caretha
;   APPLICANT: Livi, George
;   APPLICANT: Karran, Eric
;   APPLICANT: Clinkenbeard, Helen
;   APPLICANT: Browne, Michael
;   APPLICANT: Southan, Christopher
;   TITLE OF INVENTION:  HUMAN SERINE PROTEASE
;   NUMBER OF SEQUENCES:  40
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  SmithKline Beecham Corporation
;   STREET:  709 Swedeland Road
;   CITY:  King of Prussia
;   STATE:  PA
;   COUNTRY:  USA
;   ZIP:  19406
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Diskette
;   COMPUTER:  IBM Compatible
;   OPERATING SYSTEM:  DOS
;   SOFTWARE:  FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/923,454A
;   FILING DATE:
;   CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  60/025436
;   FILING DATE:  06-SEPT-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Baumeister, Kirk
;   REGISTRATION NUMBER:  33,833
;   REFERENCE/DOCKET NUMBER:  P50547
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  610-270-5096
;   TELEFAX:  610-270-5090
;   TELEX:
INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  1787 base pairs
;   TYPE:  nucleic acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
US-08-923-454A-3

Alignment Scores:
Pred. No.:
Score:                        4.39e-95
Percent Similarity:           834.50
Best Local Similarity:        60.71%
Query Match:                  42.57%
DB:                           49.70%
                              3

US-10-617-443B-2 (1-334) x US-08-923-454A-3 (1-1787)
```

```
Qy   1   SerSerProArgTyrLysPheAsnPheIleAlaAspValValGlyLysIleAlaProAla   34
         :::|||||   |:|||||||
Db  15   SerSerProArgTyrLysPheAsnPheIleAlaAspValValGlyLysIleAlaProAla   34

Qy  35   ValValHisIleGluGluPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer   54
         ||||:|||||::::|||||||
Db 259   GCTTCTCCCGGGACTGCAGTACAACTTCATCGCAGATGTGGTGGGGAAGACACCTGCC    318

Qy 319   GTGGTCTATATCGAGGATTCTGGACCGGACACCCTTTCTTGGGGCCGGAGGTCCCATACTCG   378
         :::|||||   |||::||||
Db 319   GTGGTCTATATCGAGGATTCTGGACCGGACACCCTTTCTTGGGGCCGGAGGTCCCATACTCG   378

Qy  55   SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValVal   74
         :::|||||   :|||::||
Db 379   AACGGCTCAGATTCGTCGTGTGGCTCCCGATGGGCTCCATTGTCCACCAAGCGCCCATTGGGTG   438
```

```
Qy  75   SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLeuLysValGlnLeuGlnAsnGlyAsp   94
         |||||     ---CGGCGCAGAGTCCGTGTGAGACTGCTAGGCGGGGAC   480
Db 439   GCTGAT------

Qy  95   SerTyrGluAlaThrArgIleLysTyrAsnGlyIleAspLysLysSerAspIleAlaThrIleLysIle   114
         :::|||||   :::   |||
Db 481   ACGTATGAGGCCCGGTACAGCTGTGGATCCGGATCCGCAAACGTCGAGGATT   540

Qy 115   HisProLysLysGlyLysProValLeuLeuGlnLeuHisSerAlaAspSerAspArgProGly   134
         :|||||   |||||   |||   ||||
Db 541   CAGACTAGGAGGCCTTCCCCACGGGCCGCCTGGGACCCCAGCCTGATGTCCGGCAAGGG   600

Qy 135   GluPheValValAlaLeuAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIle   154
         |||||   :::   :::|||
Db 601   GAGTTCGTTGTTGCCATGGGGAAGTCCCTTTGCACTTGCAGAACACGATCACATCCGGCATT   660

Qy 155   ValSerThrAlaGlnIleArgGlyIleArgGluLeuGlyIleLeuArgAspSerAspMetAspTyr   174
         ||||||   |||   |||:::   |||
Db 661   GTTAGCTCGTCTTCTGCTCAGCCTCCAGGCTCCAGCCAGAGACCTGGGACTCCCCAAACCAAATGTGGAATAC   720

Qy 175   IleGlnThrAspAlaIleAlaIleLeuAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeu---   193
         |||||   :::|||   |||:::|||
Db 721   ATTCAAACTGATGCAGCTATTGATTTGGAAAACTCTGGAGGTCCCCTGGTTAACCTGGTG   780

Qy 193   ----------------------   193
Db 781   AGTGAGACATCCTTCCTTCCAAGAATCCCTGCCCCAGGTCAGTGTGGGAAGGGTAGGTTT   840

Qy 193   CCCCTAATTCAAGGATGTTGGTGTCAAGTTCTTGTTGGGCTATCTCTCAA   900
Db 841   CCCCTAATTCAAGGATGTTGGTGTCAAGTTCTTGTTGGGCTATCTCTCAA   900

Qy 193   TATCCAACCAGATCTCCCACACTTGCCTGGGTACTTTTGTTCGGGTGCCCCATCCCCTA   193
Db 901   TATCCAACCAGATCTCCCACACTTGCCTGGGTACTTTTGTTCGGGTGCCCCATCCCCTA   960

Qy 194   ----------------------AspGlyGluValIleGly   199
Db 961   CTATTTGTTTAGGCTAGGGAACTGGGGGGGCTGTATCCCTGCAGGATGGGGGAGGTGATTGGA   1020

Qy 200   IleAsnThrLeuLysValThrAlaGlyIleLeuSerPheAlaIleLeuProSerAspArgIleThr   219
         :::|||   ||||||   |||
Db 1021  GTGAACAACCATGAAGGTCACAGCTGGAATCTCCTTTGCCATCCTTCTGATGGTCTTCGA   1080

Qy 220   ArgPheLeuThrGluProGlnAsp---------IleLysAspTrpLysLys   235
         |||:::   :::
Db 1081  GAGTTTCTTGCATCGTGGGGAAAAGAGAAATTCCTCTCCGGAATCAGTGGGTCCCAGCGG   1140

Qy 236   ArgPheIleGlyIleArgMetArgThrIleThrProSerLeuValAspGluLeuLysAla   255
         |||   :::   :::|||
Db 1141  CGCTACATTGGGGTGATGATGCGTACCATCACCCCGAGTCCACCACCTTGCTGGACTACAGCTT   1200

Qy 256   SerAsnProAspPheProGluValSerSerGlyIleTyrValGlnGluValAlaProAsn   275
         |||:::   :::   :::|||
Db 1201  CGAGAACCAAGCTTTCCCGAGTTCAGCAGTGGTACTGTACCAGGAAGTCATCGCTGGGGC   1260

Qy 276   SerProSerGlnAsnGlyLysIleAspGlyAspIleLeuValLysValLysValAlaGlyArg   295
         |||:::   |||:::   |||
Db 1261  TCCCCTGCCACACGGGGTCTGCGGGCGTCGTGGATCTGGTGTGAATTTGGCCATTGGGGGAGGCAG   1320

Qy 296   ProLeuGluValAspSerSerGluLeuGlnGluAlaValLeuThrGluSerProLeuLeuLeu   315
         |||:::   |||
Db 1321  ATGGTACAAAATGCTGAAGATGTTTATGAAGCTGTTGCGACCAATCCCAGTTGGCAGTG   1380

Qy 316   GluValArgArgGlyAsnAspAspLeuLeuProPheSerIleIleAlaProGluVal   332
         :::::|||   :::
Db 1381  CAGATCCGGCGGGGGGACGAGAGAAACACTGACCTTATATGTGACCCCTGAGGTC   1431
```

```
RESULT 13
US-08-923-454A-26
; Sequence 26, Application US/0892345Y
; Patent No. 6004794
; GENERAL INFORMATION:
```

```
APPLICANT:  Creasy, Caretha
APPLICANT:  Livi, George
APPLICANT:  Karran, Eric
APPLICANT:  Clinkenbeard, Helen
APPLICANT:  Browne, Michael
APPLICANT:  Southan, Christopher
TITLE OF INVENTION:  HUMAN SERINE PROTEASE
NUMBER OF SEQUENCES:  40
CORRESPONDENCE ADDRESS:
ADDRESSEE:  SmithKline Beecham Corporation
STREET:  709 Swedeland Road
CITY:  King of Prussia
STATE:  PA
COUNTRY:  USA
ZIP:  19406
COMPUTER READABLE FORM:
MEDIUM TYPE:  Diskette
COMPUTER:  IBM Compatible
OPERATING SYSTEM:  DOS
SOFTWARE:  FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/923,454A
FILING DATE:
CLASSIFICATION:  435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  60/025436
FILING DATE:  06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME:  Baumeister, Kirk
REGISTRATION NUMBER:  33,833
REFERENCE/DOCKET NUMBER:  P50547
TELECOMMUNICATION INFORMATION:
TELEPHONE:  610-270-5096
TELEFAX:  610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO:  26:
SEQUENCE CHARACTERISTICS:
LENGTH:  2551 base pairs
TYPE:  nucleic acid
STRANDEDNESS:  single
TOPOLOGY:  linear
MOLECULE TYPE:  cDNA
HYPOTHETICAL:  NO
ANTI-SENSE:  NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY:  Coding Sequence
LOCATION:  603...1733
OTHER INFORMATION:
US-08-923-454A-26

Alignment Scores:
Pred. No.:
Score:                  834.50
Percent Similarity:     60.71%
Best Local Similarity:  49.70%
Query Match:            3
DB:
```

US-10-617-443B-2 (1-334) x US-08-923-454A-26 (1-2551)

```
Qy     15  SerSerProArgTyrLysPheAsnPheIleAlaAspValValGluLysIleAlaProAla  34
               :::||||||||||||||||||||||:
Db   1023  GCTTCTCCCGGAGTACAACTTCATCGACAGTGGTGTGGAGAAGACAGCACCTGCC  1082

Qy     35  ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer  54
               :::||||||||||||||||:::||||||||::||
Db   1083  GTGGTCTATATCGAGATCCTTGGGCCCCGCGGAGGTCCCTATCCG  1142

Qy     55  SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleLeuThrAsnAlaHisValVal  74
               :::||||||||||||||||||||||||||||||||||||||
Db   1143  AACGGCTCAGGATTCGTGGTGGCTCCCGATGGGGTCATTGTTCACCAACGCCCANTGTGGTG  1202
```

<table>
<tr><td>Qy</td><td>75</td><td colspan="3">SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLeuLysValGlnLeuLeuAsnGlyAsp 94</td></tr>
<tr><td>Db</td><td>1203</td><td colspan="3">GCTGAT------------------------CGGCGGAGAGTCCGTGTGACAGTCCTAAGCGGCGAC 1244</td></tr>
<tr><td>Qy</td><td>95</td><td colspan="3">SerTyrTyrGluAlaThrIleLeuTyrAspIleAlaProGlyGlyIleGluAlaThrIleLysIle 114</td></tr>
<tr><td>Db</td><td>1245</td><td colspan="3">ACGTATGAGGCCGTGGTCACAGTCTGTGGATCCCGTGGTGCCAACGCTGAGGATT 1304</td></tr>
<tr><td>Qy</td><td>115</td><td colspan="3">HisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAlaAspLeuAsnArgProGly 134</td></tr>
<tr><td>Db</td><td>1305</td><td colspan="3">CAGACTAAGGAGGCCTCCCCACGGTGCCCTCTCGGGACAGGGCTCAGCTGATGTCCGGCAAGGG 1364</td></tr>
<tr><td>Qy</td><td>135</td><td colspan="3">GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIle 154</td></tr>
<tr><td>Db</td><td>1365</td><td colspan="3">GAGTTTGTTGTTGCCATGGGAAGTCCCTTTGCACTGCAGAACACAGTCACATCCGGCATT 1424</td></tr>
<tr><td>Qy</td><td>155</td><td colspan="3">ValSerThrArgLeuGlyIleGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyr 174</td></tr>
<tr><td>Db</td><td>1425</td><td colspan="3">GTTAGCTCTGCTCAGCCTGCCAGGGACTCCCCAAACAGTGGAATAC 1484</td></tr>
<tr><td>Qy</td><td>175</td><td colspan="3">IleGlnThrAspAlaIleIleLeuAsnThrGlyAsnSerGlyGlyProLeuValAsnLeu--- 193</td></tr>
<tr><td>Db</td><td>1485</td><td colspan="3">ATTCAAACTGATGCAGCTATTGATTTTGGAAACTCTGGAGGTCCCCTGGTTAACCTGGTG 1544</td></tr>
<tr><td>Qy</td><td>193</td><td colspan="3">-- 193</td></tr>
<tr><td>Db</td><td>1545</td><td colspan="3">AGTGAGACATCCTTCCTTCCAAGAATCCCGCCCAGGTCAGTGTGGGAAGGGTAGGTTT 1604</td></tr>
<tr><td>Qy</td><td>193</td><td colspan="3">-- 193</td></tr>
<tr><td>Db</td><td>1605</td><td colspan="3">CCCCTAATTCAAGGATGTTTGGTCAAGTTCTGAGCAGTTCTTGTTGGCTATCTCTCAA 1664</td></tr>
<tr><td>Qy</td><td>193</td><td colspan="3">-- 193</td></tr>
<tr><td>Db</td><td>1665</td><td colspan="3">TATCCAACCAGATCTCCCCAACACTTGCTGTACTTTGTTCGGGGTGCCCCCATCCCCTA 1724</td></tr>
<tr><td>Qy</td><td>194</td><td colspan="3">---------------------------AspGlyGlyValIleGly 199</td></tr>
<tr><td>Db</td><td>1725</td><td colspan="3">CTATTTGTTTAGGCTAGGGAACTGGGGGCCGCTGTATCCCTGCAGGATGGGGAGGTGATTGGA 1784</td></tr>
<tr><td>Qy</td><td>200</td><td colspan="3">IleAsnThrLeuLysValSerThrAlaGlyTyrIleSerPheAlaIleIleProSerAspArgIleThr 219</td></tr>
<tr><td>Db</td><td>1785</td><td colspan="3">GTGACACCATGAAGGTCACAGCTGGAATCTCCTTTGCCATCCCTTCTGATCGTCTTCGA 1844</td></tr>
<tr><td>Qy</td><td>220</td><td colspan="3">ArgPheLeuThrGluPheGlnAspLysLysGln------------IleLysAspTrpLysLys 235</td></tr>
<tr><td>Db</td><td>1845</td><td colspan="3">GAGTTTCTGCATCGTGGGGAAAAAGAGAATTCCTCCCCCGGGAATCAGTGGGGTCCCAGCGG 1904</td></tr>
<tr><td>Qy</td><td>236</td><td colspan="3">ArgPheIleGlyIleArgMetArgThrIleThrProSerLeuValAspGluLeuLysAla 255</td></tr>
<tr><td>Db</td><td>1905</td><td colspan="3">CGCTACATTGGGGTGCTGATGAGCTGACCCTGAGTCCCAGCATCCTTGCTGGAACTACAGCTT 1964</td></tr>
<tr><td>Qy</td><td>256</td><td colspan="3">SerAsnProAspPhePheProGluValSerSerGlyIleTyrValGlnGluValAlaProAsn 275</td></tr>
<tr><td>Db</td><td>1965</td><td colspan="3">CGAGAACCAAGCTTTCCCGATGTTCAGCATGGTGTACTGACTCATCCATAAAGTCATCCTGGGC 2024</td></tr>
<tr><td>Qy</td><td>276</td><td colspan="3">SerProSerGlnArgGlyIleGlnAspIleGlnAspIleValLysValAsnGlyArg 295</td></tr>
<tr><td>Db</td><td>2025</td><td colspan="3">TCCCCTGCCACAGGGGGCTGGTTCTGGCGTGGCTGATGTGATTTTGGCCANTGGGGAGCAG 2084</td></tr>
<tr><td>Qy</td><td>296</td><td colspan="3">ProLeuValAspSerGluLeuGlnAlaValLeuThrGluProLeuLeuLeu 315</td></tr>
<tr><td>Db</td><td>2085</td><td colspan="3">ATGGTACAAAAATGCTGAAGATGTTTATGAAGCTGTTGAACCCAAATCCCAGTTGGCAGTG 2144</td></tr>
<tr><td>Qy</td><td>316</td><td colspan="3">GluValArgArgGlyAsnAspAspLeuLeuPheSerIleAlaProGlyVal 332</td></tr>
<tr><td>Db</td><td>2145</td><td colspan="3">CAGATCCGGCGCGGGGACGAGAAAACACTGACCTTATATGTGACCCCTGAGGTC 2195</td></tr>
</table>

```
RESULT 14
US-08-923-454A-5
;  Sequence 5, Application US/08923454A
;  Patent No. 6004794
```

; GENERAL INFORMATION:
;   APPLICANT: Creasy, Caretha
;   APPLICANT: Livi, George
;   APPLICANT: Karran, Eric
;   APPLICANT: Clinkenbeard, Helen
;   APPLICANT: Browne, Michael
;   APPLICANT: Southan, Christopher
;   TITLE OF INVENTION: HUMAN SERINE PROTEASE
;   NUMBER OF SEQUENCES: 40
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SmithKline Beecham Corporation
;   STREET: 709 Swedeland Road
;   CITY: King of Prussia
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19406
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/923,454A
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/025436
;   FILING DATE: 06-SEPT-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Baumeister, Kirk
;   REGISTRATION NUMBER: 33,833
;   REFERENCE/DOCKET NUMBER: P50547
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 610-270-5096
;   TELEFAX: 610-270-5090
;   TELEX:
;
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1503 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:

US-08-923-454A-5

Alignment Scores:
Pred. No.:
Score: 3.5e-88
Percent Similarity: 779.00
Best Local Similarity: 66.87%
Query Match: 47.59%
DB: 46.40%

US-10-617-443B-2 (1-334) x US-08-923-454A-5 (1-1503)

Length: 1503
Matches: 158
Conservative: 64
Mismatches: 58
Indels: 52
Gaps: 4

Qy   15 SerSerProArgTyrLysPheAsnPheIleAlaAspValValGlyLysIleAlaProAla 34
          :::||||||:|||||||:|||||||||||||||||||||||| ||||||||||||||:||
Db   382 GCTTCTCCCCGAGTCAGTACAACTTCATCGACAGATGTGGTGGGAAGAGACAGCACCTGCC 441

Qy   35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer 54
          ::|||||||:|||||||||:||::||||||||||||||||:|:: ||:|:|||||:::|||
Db   442 GTGGTTCATATCGAGACTCCTTCTGGGCCCCGGAGGTCCCTTATCTCG 501

Qy   55 SerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValVal 74
          :::||||||||||||:||:||:|||||||||:|||||:::|||:||||:::|||||||||
Db   502 AACGGCTCAGGATTCGTCGTGTGGCTGCCGGATGGGCTCATTGTCACCAACGCCCATGGTG 561

Qy   75 SerSerAsnSerAlaAlaProGlyArgGlnLeuGlnLeuGlnAsnGlyAsp 94
          :::|||||||||||||||||||:||::::||||:|||||:||||:|||||
Db   562 GCTGAT---------------CGGCGCAGAGTCCGTGTGAGACTGTAAGCGGCGGAC 603

Db   562 GCTGAT---------------CGGCGCAGAGTCCGTGTGAGACTGTAAGCGGCGGAC 603

Qy   95 SerTyrGluAlaThrIleLeuTyrAspIleAspLysLysSerAspIleAlaThrIleLysIle 114
          :::||||:|||||:|:|||||||||:|:|||:||||:|||||:||:::|||:||:|||||
Db   604 AGTTATGAGGCCGGTGGTCACGAGCTGTGGATCCGGGCCAGACGATCGGAACGGCTGGGATT 663

Qy   115 HisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAlaAspLeuArgProGly 134
          ||::||:|||||||||||||||:|:|||||||||||||:::|:|||||::::||:||||
Db   664 CAGACTAAGGAGCCTCTCCCCAACGGTCGCCTCTGGGACGGCTCAGCTGATGTCCGGCAAGG 723

Qy   135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIle 154
          ||:||:||:|:||||||||||:|||||||||||||||||||||:::|:||||:||:|||
Db   724 GAGTTGTTGTTGCCATAGGGAGCCCTTTCGCACTGCAGAACACAGTCACCACCGGCATT 783

Qy   155 ValSerThrAlaGlnHisArgGlyGluGlyLeuGlyLeuArgAspSerAspMetAspTyr 174
          :::||:|::||||||||||:||:|||:|||||||:||:|||||:|:||:|::||||||
Db   784 GTTAGCTCTGCTCAGCGTCACCAGCGAGGACTCCCCAAACCAAGTGGAATAC 843

Qy   175 IleGlnThrAspAlaIleIleIleLeuAsnThrTyrGlyAspGlyGlyProLeuValAsnLeu--- 193
          ||::|||:|||:||:::|||||::|||:||::|:||:|||||||:::||:||:||::
Db   844 ATTCAAACTGATGCAGTTATTGATTTTGGAAACTCTGGAGGTCCCTTAACCTGCT 903

Qy   194 ------------------AspGlyValIleIleGlyIleAsnThrLeuLys 204
                            ::||:::||||||:::|||||||||::|||||
Db   904 AGGGACTGGGGGTCTGTATCCCTGCAGGATGGGGGTGAGGTGAACACCATGAAG 963

Qy   205 ValThrAlaGlyIleSerPheAlaIleProSerAspAspArgIleThrArgPheLeuValThrGlu 224
          :::||||||||:::|||||||||:::|||::||::||:||||:::|||||:|||:|::||
Db   964 GTCACAGCTGGAATCTCCTTTGCCATCCCTTCTGATGGTCGTATCACTCGT 1023

Qy   225 PheGlnAspLysLysGln--------IleLysLysAspTrpLysLysArgPheIleGlyIle 240
          |||||:||:|||||||| ::|||:::||||||:::||:||||||:|||||
Db 1024 GGGGAAAAAGAAGAAATTCCTCCTCCGGAATCAGTGGGTCCCAGCGGCGCCCTACATTGGGGTG 1083

Qy   241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysValAlaSerAsnProAspPhe 260
          |||:::|||:::|||:|||||||||||||||||||||||:||:::||||||:||||||:::
Db 1084 ATGATGCTGACCCTGAGTCCC---------------- 1104

Qy   261 ProGluValSerSerGlyIleTyrValGlnGluValLeuAlaAlaProAsnSerProSerGlnArg 280
          ||:|||:::::||||||||||||||:|::|:||||:||:::||:::||:::|||:::::||
Db 1105 ----------------AGG 1107

Qy   281 GlyGlyIleGluGlyAsnGlyIleIleValLysValAlaAsnGlyArgProLeuValAspSer 300
          |||||:|:|::||||||::|||::|||:||:||:||||:|||:::||:::||::::
Db 1108 GCTGGTCTGCGGCGCCCTGGTGAGGTTGATTTTGGCCATTGGAGCAGGATGGGTACAAAATGCT 1167

Qy   301 SerGluLeuGlnGluValAlaValLeuThrGluSerProLeuLeuLeuGluValArgGly 320
          ||::|||::|||||:::||:|||||:||:||:||||||||||||||||||:::||:|||
Db 1168 GAAGATGTTTATGAAGGTGTTCGAACCCCAATCCCAGTTGGCAGTTGCAGAGTCCGGCGGGA 1227

Qy   321 AsnAspAspLeuLeuPheSerIleIleAlaProGluVal 332
          :::||||||::||||:::||:::|:||||||:||||||
Db 1228 CGAGAAACACTGACCTTATATGTGACCCCTGAGGTC 1263

RESULT 15
US-08-923-454A-28
; Sequence 28, Application US/0892345A
; Patent No. 6004794
; GENERAL INFORMATION:
;   APPLICANT: Creasy, Caretha
;   APPLICANT: Livi, George
;   APPLICANT: Karran, Eric
;   APPLICANT: Clinkenbeard, Helen
;   APPLICANT: Browne, Michael
;   APPLICANT: Southan, Christopher
;   TITLE OF INVENTION: HUMAN SERINE PROTEASE
;   NUMBER OF SEQUENCES: 40
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SmithKline Beecham Corporation
;   STREET: 709 Swedeland Road
;   CITY: King of Prussia
;   STATE: PA
;   COUNTRY: USA

```
;                ZIP:  19406
; COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Diskette
;   COMPUTER:  IBM Compatible
;   OPERATING SYSTEM:  DOS
;   SOFTWARE:  FastSEQ Version 1.5
; CURRENT APPLICATION INFORMATION:
;   APPLICATION NUMBER:  US/08/923,454A
;   FILING DATE:
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  435
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  60/025436
;   FILING DATE:  06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME:  Baumeister, Kirk
;   REGISTRATION NUMBER:  33,833
;   REFERENCE/DOCKET NUMBER:  P50547
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  610-270-5096
;   TELEFAX:  610-270-5090
;   TELEX:
; INFORMATION FOR SEQ ID NO:  28:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2144 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
;   FEATURE:
;     NAME/KEY:  Coding Sequence
;     LOCATION:  603...1910
;   OTHER INFORMATION:

US-08-923-454A-28

Alignment Scores:
Pred. No.:
Score:                    6.33e-88
                          779.00
Percent Similarity:       66.87%
Best Local Similarity:    47.59%
Query Match:              46.40%
DB:                       3

US-10-617-443B-2 (1-334) x US-08-923-454A-28 (1-2144)

                                                              Length:         2144
                                                              Matches:         158
                                                              Conservative:     64
                                                              Mismatches:       58
                                                              Indels:           52
                                                              Gaps:              4
```

```
Qy     15 SerSerProArgTyrLysPheAsnPheIleAlaAspValValGluLysIleAlaProAlaAla 34
          :::|||||||||          ::::|||           |||       |||:::|||
Db   1023 GCTTCTCCCCGGAGTCAGTACAACTTCATCCGCGATGTGGTCGGAGAAGACACGCCACCTGCC 1082

Qy     35 ValValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSer 54
          |||:::||| |||||||:::           :::||||||||||        |||:::
Db   1083 GTGGTCTATATCGGAGATCCTGGACACGGCACCCTTCTTTGGGCGCCGAGGTCCCTATCTCG 1142

Qy     55 SerGlySerGlyPheIleMetSerGluLysAlaGlyLysIleLeuIleThrAsnAlaHisVal 74
          :::|||||||||||  |::::::::::
Db   1143 AACGGCTCAGGATTCGTGGCTGGCCTGCCGATGGGGCTCATTGTCACCAACGCCCATGTGGTG 1202

Qy     75 SerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGlnLeuAsnGlyAsp 94
          |||||||||||                           :::||||||::: |||
Db   1203 GCTGAT------------------CGGCGCAGAGTCCCGTGAGACTTGCTAAGGCGGGCGAC 1244

Qy     95 SerTyrGluAlaThrIleLysAspIleLysSerAspIleAlaThrIleLysIle 114
          :::|||||||||           ::: |||               :::||||||
Db   1245 ACGTATGAGGCCCTGGTGCACAGCTGGGATCCCGCAGACATCGCAACGCTGAGGATT 1304

Qy    115 HisProLysLysLysLeuProValLeuLeuLeuHisSerAlaAspLeuArgProGly 134
          ||:::||||||||                             :::|||
Db   1305 CAGACTAAGGAGCCTCTCCCCACGCTGCCTCTGGGACCGCTCAGCTGATTGTCCCGCAAGG 1364

Qy    135 GluPheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIle 154
          |||:::||| |||||||       :::|||           |||
```

```
Db   1365 GAGTTTGTTGTTGCCATCGGAAGTCCCTTTGCACTGCAGAACACGGTCACCACCGGCATT 1424

Qy    155 ValSerThrAlaGlnAsnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyr 174
          ::: ||||||             :::|||                          :::||||||
Db   1425 GTTAGCTCTGCTCAGCGTCCAGGAGACCTGGGACTCCCCCAAACCAATGTGGAATAC 1484

Qy    175 IleGlnThrAspAlaIleIleLeuAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeu-- 193
          ||||||       ::: |||                   |||||||||:::
Db   1485 ATTCAAACTGATGCAGCTATTGATTTTGGAAACTTCGGAGGTCCCCTGGTTAACCTGGCT 1544

Qy    194 --------------------AspGlyValIleGlyIleAsnThrLeuLys 204
                              |||       ::: |||||||
Db   1545 AGGGGACTGGGGGGCGGTGTATCCCTGGAGGTGAGGTGAGTGGAGTGGAACACCATGAAG 1604

Qy    205 ValThrAlaGlyValIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGlu 224
          |||||||:::                ||||||:::|||           |||||||
Db   1605 GTCACAGCTGGAGTCATCCCTTTGCCATCCCTTCTGATCGTCTTCGAGAGTTTCTGCATCGT 1664

Qy    225 PheGlnAspLysGln---------IleLysAspTrpLysTyrLysAspPheIleGlyIle 240
          ::: |||                 :::||||||:::|||:::
Db   1665 GGGGAAAAGAGAAATTCCTCCTCCGGAATCAGTGGGTCCCAGGCGGCGCTACATTGGGGTG 1724

Qy    241 ArgMetArgThrIleThrProSerLeuValAspGluLysAlaSerAsnProAspPhe 260
          |||         ::: |||||||       :::
Db   1725 ATGATGCTTGACCCTGAGTCCC------------------------------------ 1745

Qy    261 ProGluValSerSerGlyIleThrTyrValGlnValAlaProAsnProSerGlnArg 280
                                                            |||
Db   1746 ------------------------------------------------------AGG 1748

Qy    281 GlyGlyIleGluGlyAsnGlyAspIleLeuValLysValAlaSerGlyTyrProLeuGluSer 300
          |||:::|||           :::||||||     :::           |||:::
Db   1749 GCTGGTCTTGCGCGCCTGGTTGATGGTGATTTTGGCCATTGGGAGGACAGAATGGGTACAAATGCT 1808

Qy    301 SerGluLeuGlnAlaValLeuThrGluSerProLeuLeuLeuGluValArgArgGly 320
          :::::::           |||:::||| :::::::       |||
Db   1809 GAAGATGTTTATGAAGCTGTTCGAACCCAAATCCCAGTTGGCAGTGCAGAATCCGGCGGGCA 1868

Qy    321 AsnAspAspLeuLeuPheSerIleAlaProGluVal 332
          :::|||           |||
Db   1869 CGAGAAACACTGACCTTATATGTGACCCCTGAGGTC 1904
```

```
Search completed: June 30, 2005, 01:32:59
Job time : 320.971 secs
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 21:23:47 ; Search time 1045.82 Seconds
                (without alignments)
                1996.208 Million cell updates/sec

Title: US-10-617-443B-2
Perfect score: 1679
Sequence: 1 MHLALPASAGLHQLSSPRYK.........LEVRGNDDLLFSIAPEVVM 334

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0 , Fgapext 7.0
               Delop 6.0 , Delext 7.0

Searched: 6067389 seqs, 3125258755 residues

Total number of hits satisfying chosen parameters: 12134778

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10617443/runat_29062005_135933_26126/app_query.fasta_1.718
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10617443 @CGN 1 1 582 @runat_29062005_135933_26126
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100

Database : Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10I_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US50_PUBCOMB.seq:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1679	100.0	3006	17	US-10-189-099A-1	Sequence 1, Appli
2	1679	100.0	3006	21	US-10-617-443B-1	Sequence 1, Appli
3	1635	97.4	1044	9	US-09-796-858-9	Sequence 9, Appl
4	1635	97.4	1563	10	US-09-796-753-33	Sequence 33, Appl
5	1635	97.4	1797	18	US-10-275-505-28	Sequence 28, Appl
6	1635	97.4	2040	9	US-09-818-143-14	Sequence 14, Appl
7	1635	97.4	2094	19	US-10-381-820A-7	Sequence 7, Appli
8	1635	97.4	2541	21	US-10-956-157-5181	Sequence 5181, Ap
9	1635	97.4	2543	21	US-10-485-313A-31	Sequence 31, Appl
10	1635	97.4	2576	10	US-09-796-753-31	Sequence 31, Appl
11	1536	91.5	2450	21	US-10-485-313A-26	Sequence 26, Appl
12	1482.5	88.3	2551	15	US-10-301-822-78	Sequence 78, Appl
13	1378	82.1	1868	9	US-09-764-898-112	Sequence 112, App
14	1122	66.8	1695	9	US-09-764-898-112	Sequence 112, Appli
15	1117	66.5	1953	21	US-10-485-313A-32	Sequence 32, Appl
16	1105	65.8	1788	9	US-09-764-898-40	Sequence 40, Appl
17	1057	63.0	1897	21	US-09-485-313A-38	Sequence 38, Appl
18	1034.5	61.6	1440	9	US-09-796-858-7	Sequence 7, Appli
19	1034.5	61.6	1894	17	US-10-104-047-795	Sequence 795, App
20	1034.5	61.6	2034	15	US-10-102-524-1732	Sequence 1732, Ap
21	1034.5	61.6	2036	10	US-09-706-516	Sequence 516, App
22	1034.5	61.6	2036	17	US-10-170-385-356	Sequence 356, App
23	1034.5	61.6	2036	17	US-10-172-118-815	Sequence 815, App
24	1034.5	61.6	2036	18	US-10-342-887-815	Sequence 815, App
25	1034.5	61.6	2036	21	US-10-772-636-47	Sequence 47, Appl
26	1034.5	61.6	2036	21	US-10-956-157-919	Sequence 919, App
27	1034.5	61.6	2054	14	US-10-084-817-32	Sequence 32, Appl
28	1034.5	61.2	2437	14	US-10-198-846-10826	Sequence 10826, A
29	1027.5	61.2	2157	17	US-10-264-004-135	Sequence 135, App
30	1006.5	59.9	1855	9	US-09-925-298-134	Sequence 134, App
31	1006.5	59.9	1855	14	US-10-102-806-134	Sequence 134, App
32	1006.5	59.9	2212	20	US-10-723-860-5299	Sequence 5299, Ap
33	1005.5	59.9	1534	9	US-09-765-231A-20	Sequence 20, Appl
34	981.5	58.5	1534	9	US-10-764-420-1506	Sequence 1506, Ap
35	945	56.3	1328	21	US-10-485-231A-20	Sequence 20, Appl
36	920	54.8	1859	10	US-09-866-050A-588	Sequence 588, App
37	920	54.8	1859	14	US-10-152-661-588	Sequence 588, App
38	890	53.0	1544	19	US-10-789-241-49	Sequence 49, Appl
39	890	53.0	1544	21	US-10-772-636-7	Sequence 7, Appl
40	882	52.5	963	19	US-09-935-390A-18	Sequence 18, Appl
41	882	52.5	975	19	US-10-730-476A-11	Sequence 11, Appl
42	882	52.5	1374	14	US-10-197-634-14	Sequence 14, Appl
43	882	52.5	1400	21	US-10-956-157-7054	Sequence 7054, Ap
44	882	52.5	1597	17	US-10-352-684A-9	Sequence 9, Appli
45	882	52.5	1597	21	US-10-956-157-1819	Sequence 1819, Ap

RESULT 1
US-10-189-099A-1
; Sequence 1, Application US/10189099A
; Publication No. US20040005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1

ALIGNMENTS

```
;  LENGTH: 3006
;  TYPE: cDNA
;  ORGANISM: Homo sapiens
US-10-189-099A-1
US-10-617-443B-2

Alignment Scores:
Pred. No.:
Score:                      2.43e-180
Matches:                    1679.00
Percent Similarity:         100.00%        Conservative:       334
Best Local Similarity:      100.00%        Mismatches:         0
Query Match:                17             Indels:             0
DB:                                        Gaps:               0

US-10-617-443B-2 (1-334) x US-10-189-099A-1 (1-3006)

Qy      1  MetHisLeuAlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLys  20
Db   1011  ATGCACCTGGCCCTTCCCGCCAGCGCAGGTCTCCACCAGCTGAGCAGCCCGCGCTACAAG  1070

Qy     21  PheAsnPheIleAlaAspValValGlyLysIleAlaProAlaValValHisIleGluLeu  40
Db   1071  TTCAACTTCATTGCTGACGTGGTGGGAAAGATCGCACCAGCCGTGGTCCACATAGAGCTC  1130

Qy     41  PheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIle  60
Db   1131  TTCCTGAGACACCCGCTGTTTGGCCGAAACGTGCCCCTGTCCAGCGGGTTCTGGCTTCATC  1190

Qy     61  MetSerGluValGlyLeuIleThrAsnAlaHisValSerSerAsnSerAlaAla  80
Db   1191  ATGTCAGAGGCCGGCCTGATCACCAATGCCCACGTGTCGTCCAGCAACAGTGCTGCC  1250

Qy     81  ProGlyArgGlnIleLeuLeuValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIle  100
Db   1251  CCGGGCAGGCAGATCCTCAAGGTGCAGCTACAGAATGGGGACTCCTATGAGGCCACCATC  1310

Qy    101  LysAspIleAspLysLysSerAspIleAlaThrIleLysIleHisProLysLysLysLeu  120
Db   1311  AAAGACATCGACAAGAAGAGTCGACATTGCCACCATCAAGATCCATCCCAAGAAAAAGTC  1370

Qy    121  ProValLeuLeuLeuGlyHisSerAlaAlaAspLeuArgProGlyGluPheValAlaIle  140
Db   1371  CCTGTGTTGCTGGTCACTCGGCCCCGGCCTGGGGAGTTTGTGGCGATC  1430

Qy    141  GlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArg  160
Db   1431  GGCAGTCCCTTCGCCCTACAGAACACAGTGACAACGGGCCATCTGCACGGCACTCCAGCGG  1490

Qy    161  GluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrrIleGlnThrAspAlaIle  180
Db   1491  GAGGGCAGGGAGCTGGGCCTCCGGGACTCCGACATGGACTACATCCAGACGGATGCCATC  1550

Qy    181  IleAsnTyrGlyValAsnSerGlyGlyProLeuValLysAspSerGluValLeuGlyIle  200
Db   1551  ATCAACTACGGGGTAAACTCCGGGGGACCACTGGTGAAGGACTCGGAGGTCATTGGCATC  1610

Qy    201  AsnThrLeuLysValIleThrLeuAlaGlyIleSerProPheAlaIleProArgSerAspArgIleThrArg  220
Db   1611  AACACGCTCAAGGTCACGGCTGCCATCCCCTCAGACCGCATCACGCGG  1670

Qy    221  PheLeuThrGlnPheGlnAspIleAlaSerAsnProAspPhe  240
Db   1671  TTCCTCACAGAGTTCCAAGACACAGCAGATCAAAGACACTGAAGGCCGTTCATCGGCATA  1730

Qy    241  ArgMetArgThrIleThrProSerLeuValAspArgGluLeuLysAlaSerAsnProAspPhe  260
Db   1731  CGGATGCGGACGGATCACACCAACCCTGGTGATGAGCTGAAAGGCCAGGAACCCGGACTTC  1790

Qy    261  ProGluValSerSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGlnArg  280
Db   1791  CCAGAGGTCAGCAGTGGAATTTATGTCCAAGAGGTTGCGCCCAATTCACCTTCTCAGAGA  1850

Qy    281  GlyGlyIleGluGlyAspGlyAspProIleIleValAlaLysAsnGlyArgProLeuValAspSer  300
Db   1851  GGGGGCCATCCAAGAGATGGTGGACATCATCGTTGCACAAGGTGTTCCTTAGTGGGGACTCG  1910

Qy    301  SerGluLeuGlnGluValAlaLeuThrGluSerProLeuLeuGluGlyValValArgArgGly  320
Db   1911  AGTGAGCCTGCAGGAGGCCGGTGCTACAGCGAGTCCCTCCTCTACTGGAGGGGTGCGTGGGG  1970

Qy    321  AsnAspAspProLeuPheSerIleIleAlaProGlyGluValValMet  334
Db   1971  AACGACGACCCTCCTCTTCAGCATCGCACCTGAGGTGGTCATG  2012

RESULT 2
US-10-617-443B-1
; Sequence 1, Application US/10617443B
; Publication No US20050019777A1
; GENERAL INFORMATION:
;  APPLICANT: Darrow, Andrew L
;  APPLICANT: Qi, Jian-shen
;  APPLICANT: Chen, Cailin
;  APPLICANT: Andrade-Gordon, Patricia
;  TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
;  FILE REFERENCE: ORT-1644
;  CURRENT APPLICATION NUMBER: US/10/617,443B
;  CURRENT FILING DATE: 2003-07-02
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
;  LENGTH: 3006
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-617-443B-1

Alignment Scores:
Pred. No.:
Score:                      2.43e-180
Matches:                    1679.00
Percent Similarity:         100.00%        Conservative:       334
Best Local Similarity:      100.00%        Mismatches:         0
Query Match:                21             Indels:             0
DB:                                        Gaps:               0

US-10-617-443B-2 (1-334) x US-10-617-443B-1 (1-3006)

Qy      1  MetHisLeuAlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLys  20
Db   1011  ATGCACCTGGCCCTTCCCGCCAGCGCAGGTCTCCACCAGCTGAGCAGCCCGCGCTACAAG  1070

Qy     21  PheAsnPheIleAlaAspValValGlyLysIleAlaProAlaValValHisIleGluLeu  40
Db   1071  TTCAACTTCATTGCTGACGTGGTGGGAAAGATCGCACCAGCCGTGGTCCACATAGAGCTC  1130

Qy     41  PheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIle  60
Db   1131  TTCCTGAGACACCCGCTGTTTGGCCGAAACGTGCCCCTGTCCAGCGGGTTCTGGCTTCATC  1190

Qy     61  MetSerGluValGlyLeuIleThrAsnAlaHisValSerSerAsnSerAlaAla  80
Db   1191  ATGTCAGAGGCCGGCCTGATCACCAATGCCCACGTGTCGTCCAGCAACAGTGCTGCC  1250

Qy     81  ProGlyArgGlnIleLeuLeuValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIle  100
Db   1251  CCGGGCAGGCAGATCCTCAAGGTGCAGCTACAGAATGGGGACTCCTATGAGGCCACCATC  1310

Qy    101  LysAspIleAspLysLysSerAspIleAlaThrIleLysIleHisProLysLysLysLeu  120
Db   1311  AAAGACATCGACAAGAAGAGTCGACATTGCCACCATCAAGATCCATCCCAAGAAAAAGTC  1370

Qy    121  ProValLeuLeuLeuGlyHisSerAlaAlaAspLeuArgProGlyGluPheValAlaIle  140
Db   1371  CCTGTGTTGCTGGTCACTCGGCCCCGGCCTGGGGAGTTTGTGGCCATC  1430

Qy    141  GlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArg  160
Db   1431  GGCAGTCCCTTCGCCCTACAGAACACAGTGACAACGGGCCATCTGCACGGCACTCCAGCGG  1490
```

```
Qy    161  GluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIle  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491  GAGGGCCAGGGAGCTGGGCGCCTCCGGGGACTCCGACATGGACTACATCCAGACGGATGCCATC 1550

Qy    181  IleAsnThrTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIle  200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551  ATCAACTACGGGAACTCCGGGGGACCACTGGTGAACCTGGATGGGGAGGTCATTGGCATC 1610

Qy    201  AsnThrLeuLysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArg  220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611  AACACGCTCAAGGTCACGGCTGGGATCTCCTTTGCCATCCCCTCAGACCGCATCACACGG 1670

Qy    221  PheLeuThrGluPheGlnAspLysAsnGlnIleIleLysLysArgPheIleGlyIle  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1671  TTCCTCACAGAGTTCCAAGACAAGAACCAGATCAAAAGACTGGAAGAAGCGCTTCATCGGCATA 1730

Qy    241  ArgMetArgThrIleThrProSerLeuValAspSerGluLeuLysLysAlaSerAsnProAsnPhe  260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1731  CGGATGCGGACGATCACACCAAGCCTGGTGGATGACTGAGGCCAAGAAGGCCAGCAACCCGGACTTC 1790

Qy    261  ProGluValSerSerGlyIleTyrValGlnGluValAlaAlaProAsnSerProSerGlnArg  280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1791  CCAGAGGTCAGCAGTGGAATTTACGTGCAAGAAGTTGCGCCGAATTCACCTTCCAGAGA 1850

Qy    281  GlyGlyValIleGlnAspGlyAspIleIleValLysSerValAsnGlyArgProLeuValAspSer  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1851  GGCGGCGTCATCCAAGATGGTGACATCATCGTCAAGGTCAACGGCCGTCCTCTAGTGGACTCG 1910

Qy    301  SerGluLeuGlnLeuAlaValLeuThrGluSerProLeuLeuLeuGluLeuValArgArgGly  320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1911  AGTGAGCTGCAGAGGCCGTGCTGACCGAGTCTCTCTCTACTGGAGGTGGTGCGGCGGGG 1970

Qy    321  AsnAspAspLeuLeuPheSerIleLeuAlaProGluValValMet  334
           |||||||||||||||||||||||||||||||||||||
Db   1971  AACGACGACCTCCTTCTTCAGCATCCGACACCTGAGGGTGGTCATG 2012
```

```
Alignment Scores:
Pred. No.:                      5.42e-176    Length:              1044
Score:                          1635.00      Matches:             325
Percent Similarity:             100.00%       Conservative:       0
Best Local Similarity:          100.00%       Mismatches:         0
Query Match:                    97.38%        Indels:             0
DB:                             9             Gaps:               0
```

US-10-617-443b-2 (1-334) x US-09-796-858-9 (1-1044)

```
Qy     10  GlyLeuHisGlnLeuSerSerProArgThrLysPheAsnPheIleAlaAspValValGlu  29
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10  GGTCTCCACCAGCTGAGCAGCCCGCGTACGAAGTTCAACTTCATTGCTGACGTGGTGGAG  69

Qy     30  LysIleIleAlaProAlaAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg  49
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70  AAGATCGCCCACCAGCCGCTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC  129

Qy     50  AsnValProLeuSerSerGlySerGlyLysPheProPheIleMetSerGluValAlaGlyLysPro  69
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130  AACGTGCCCCTGTCCAGCGGTTCTGGCTTCATCATGTCAGAGGTCGCGGCCTGATCATCACC  189

Qy     70  AsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnGlnLeuLeuLysLysGln  89
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190  AATGCCCACGTGGTGTCCAGCAACACTGCTGCCGCCCCGGGCAGGCAGCAGCTCAAGGTGCAG  249

Qy     90  LeuGlnAsnGlyAspSerTyrGluAlaThrArgThrLeuLysProValLeuGlyHisSerAla  109
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250  CTACAGAATGGGGACTCCTATGAGGCCACCAGAAAAAGCTCCCTGTGTTGTTGGCTGGTCACTCGGCC  309

Qy    110  AlaThrIleLeuIleGluHisProLeuLysLysGlyLeuProValLeuLeuLeuGlyHisSerAla  129
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310  GCCACCATCAAGATCATCCAAGAAACAGTGCTCGTCCCGGGGCCAGGCAGCAGCTCAAGGTGCAG  369

Qy    130  AspLeuArgProGlyGluPheValValAlaIleAlaIleGlySerProPheAlaLeuGlnAsnThr  149
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370  GACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGCCATCGGCAGTCCCTTCGCCTACGAGAACACA  429

Qy    150  ValThrThrGlyIleValSerThrArgGlnArgGlyGluGlyLeuArgAsp  169
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430  GTGACAACGGGCATCGTCAGCACTCGCCCAGCGGGAGGGCACGGGAGCGTCCCGGGAC  489

Qy    170  SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnThrTyrGlyAsnSerGlyGlyPro  189
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490  TCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGACCA  549

Qy    190  LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle  209
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550  CTGGTGAACCTGGATGGGGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGGATCTC  609

Qy    210  SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysLysGln  229
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610  TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGAAGCAG  669

Qy    230  IleLysLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu  249
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    670  ATCAAAAGACTGGAAGAAGCGCTTCATCGGCATACGGATGCGGACGATCACACCAAGCCTG  729
```

```
Qy    250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleIleTyrrVal 269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    790 GTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTCAGCAGTGGAATTTATGTG 849

Qy    270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGluGlnAspGlyIleIle 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    850 CAAGAGGTTGCGCCGCAAATCACCTTCTCAGAGAGGGCCATCCAAGATGGTGACATCATC 909

Qy    290 ValLysValAsnGlyArgProLeuValAlaSerGluLeuGlnLeuAlaValLeuThr 309
          ||||||||||||||||||||||||||||||||||||
Db    910 GTCAAGGTCAACGGGCGGCGTCCTCTAGTGGCATCGGAGCTGCAGAGGCCGTGCTGACC 969

Qy    310 GluSerProLeuLeuLeuGluValArgGlyAsnAspAspLeuLeuPheSerIleAla 329
          ||||||||||||||||||||||||||||||||||||||
Db    970 GAGTCTCCTCTCCTACTGGAGGTGCGGGGCGGGGGACGACCACTCCTCCTTCAGCATCGCA 1029

Qy    330 ProGluValValValMet 334
          ||||||||||||||
Db   1030 CCTGAGGTGGTCATG 1044
```

RESULT 4
US-09-796-753-33
Sequence 33, Application US/09796753
Publication No. US20030027998A1
GENERAL INFORMATION:
  APPLICANT: McCarthy, Sean A.
  TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
  FILE REFERENCE: 7853-227-999
  CURRENT APPLICATION NUMBER: US/09/796,753
  CURRENT FILING DATE: 2001-03-01
  PRIOR APPLICATION NUMBER: 09/183,175
  PRIOR FILING DATE: 1998-10-30
  PRIOR APPLICATION NUMBER: 09/223,094
  PRIOR FILING DATE: 1998-12-30
  PRIOR APPLICATION NUMBER: 09/223,546
  PRIOR FILING DATE: 1998-12-30
  PRIOR APPLICATION NUMBER: 09/224,246
  PRIOR FILING DATE: 1998-12-30
  PRIOR APPLICATION NUMBER: 09/259,388
  PRIOR FILING DATE: 1999-02-26
  PRIOR APPLICATION NUMBER: 60/122,458
  PRIOR FILING DATE: 1999-03-01
  PRIOR APPLICATION NUMBER: 09/312,359
  PRIOR FILING DATE: 1999-05-14
  PRIOR APPLICATION NUMBER: 09/342,687
  PRIOR FILING DATE: 1999-06-29
  PRIOR APPLICATION NUMBER: 09/336,536
  PRIOR FILING DATE: 1999-06-18
  PRIOR APPLICATION NUMBER: 09/345,464
  PRIOR FILING DATE: 1999-06-30
  PRIOR APPLICATION NUMBER: 09/365,164
  PRIOR FILING DATE: 1999-07-30
  PRIOR APPLICATION NUMBER: 09/399,723
  PRIOR FILING DATE: 1999-09-20
  PRIOR APPLICATION NUMBER: 09/409,634
  PRIOR FILING DATE: 1999-09-30
  PRIOR APPLICATION NUMBER: 09/471,179
  PRIOR FILING DATE: 1999-12-23
  PRIOR APPLICATION NUMBER: 09/474,072
  PRIOR FILING DATE: 1999-12-29
  PRIOR APPLICATION NUMBER: 09/474,071
  PRIOR FILING DATE: 1999-12-29
  PRIOR APPLICATION NUMBER: 09/514,010
  PRIOR FILING DATE: 2000-02-25
  PRIOR APPLICATION NUMBER: 09/516,745
  PRIOR FILING DATE: 2000-03-01
  PRIOR APPLICATION NUMBER: 09/572,002
  PRIOR FILING DATE: 2000-05-14
  PRIOR APPLICATION NUMBER: 09/597,993
  PRIOR FILING DATE: 2000-06-19
  PRIOR APPLICATION NUMBER: 09/599,596
  PRIOR FILING DATE: 2000-06-22
  PRIOR APPLICATION NUMBER: 09/630,334
  PRIOR FILING DATE: 2000-07-31
  PRIOR APPLICATION NUMBER: 09/606,565
  PRIOR FILING DATE: 2000-06-29
  PRIOR APPLICATION NUMBER: 09/606,317
  PRIOR APPLICATION NUMBER: 09/665,666
  PRIOR FILING DATE: 2000-09-20
  PRIOR APPLICATION NUMBER: 09/677,751
  PRIOR FILING DATE: 2000-09-30
  NUMBER OF SEQ ID NOS: 162
SEQ ID NO 33
  LENGTH: 1563
  TYPE: DNA
  ORGANISM: Mouse
US-09-796-753-33

Alignment Scores:
  Pred. No.:
  Score: 9.72e-176   Length: 1563
  Percent Similarity: 1635.00   Matches: 325
  Best Local Similarity: 100.00%   Conservative: 0
  Query Match: 100.00%   Mismatches: 0
                 97.38%   Indels: 0
  DB: 10   Gaps: 0
```

US-10-617-443B-2 (1-334) x US-09-796-753-33 (1-1563)

```
Qy     10 GlyLeuHisGlnLeuSerSerProArgThrLysPheAsnPheIleAlaAspValValGlu 29
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 GGTCTCCACCAGCTGAGCAGCCCGCGGACAAAGTTCAACTTCATTGCTGACGTGGTGGAG 396

Qy     30 LysIleAlaProAlaValValHisIleGluLeuProLeuPheLeuArgHisProLeuPheLeuArg 49
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACCACCCCTGTTTGGCCGC 456

Qy     50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGlyGluValAlaGlyIleLeuIleLeuThr 69
          |||||||||||||||||||||||||||||||||||||||||
Db    457 AACGTGCCCCTGTCCAGCGGGTCTCGGCTTCATCATGTCAGGAGGGCGCGGGCCCTGANCATCACC 516

Qy     70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyGlnLeuLeuLysValGln 89
          ||||||||||||||||||||||||||||||||||||
Db    517 AATGCCCACGTGGTGTCCAGCAACACAGTGCTGCCCCCGGGCCAGCAGCTCAAGGTGCAG 576

Qy     90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysIleGlyGlyLysGlySerAspIle 109
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    577 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAATCGGAGGAAAAGGCTCAGACATT 636

Qy    110 AlaThrIleLysIleGlnHisIleSerProLysLysGlyLeuProValLeuLeuLeuGlyHisSerAla 129
          ||||||||||||||||||||||||||||||||||||||||||
Db    637 GCCACCATCAAGATCCAACCCAAAGAAAAAGGTCCCCGTGTTGTTGCTGGGCACTCGGCC 696

Qy    130 AspLeuArgProGlyPheAlaIleAlaIleAlaIleAsnThrThr 149
          ||||||||||||||||||||||||||||||||||||
Db    697 GACCTGCGCCGGGACTTTGTGGTCCACATCGGACGTCCCCTTCGCCCTACACAGAACACA 756

Qy    150 ValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyLeuArgAsp 169
          ||||||||||||||||||||||||||||||||
Db    757 GTGACACAACGGGCCATCGTCAGCACTGCCCAGAGCGGCAGGGAGGGCCTTCGGGAC 816

Qy    170 SerAspMetAspTyrIleGlnThrArgAspAlaIleIleIleAsnThrLeuAsnSerGlyGlyPro 189
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    817 TCCGACATGGACTACATCCAGACGCGAGATGCCATCAACACGCTGAACTCCGGGGGACCA 876

Qy    190 LeuValAsnLeuAspGlyValIleGlyValThrAlaGlyIle 209
          ||||||||||||||||||||||||||||||||
Db    877 CTGGTGAACCTGGATGGCGGAGTCATTGGCATCAACACGCTCAAGGTCACCGCGGCATC 936

Qy    210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluGlnGluLysAspGln 229
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCAAGAACAACCAG 996

Qy    230 IleLysAspTrpLysLysArgPheIleGluArgMetArgThrIleThrProSerLeu 249
          ||||||||||||||||||||||||||||||||||||||
Db    997 ATCAAAGACTGGAAGAAGCGCTTCATCGGAGCATACGGATGCCGACGATCACACCAAGCCTG 1056
```

```
Qy    250  ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057  GTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTCAGCAGTGGGAATTTATGTG 1116

Qy    270  GlnGluValAlaProAsnSerProGlnArgGlyIleGlnAspGlyIleIleTyrIle 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1117  CAAGAGGTTGCGCCGAATTCACCTTCCACGAGGAGGCCATCCAAGATGGGTGACATCATC 1176

Qy    290  ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnLeuAlaValLeuThr 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1177  GTCAAGGTCAACGGGCGGCCGGTCCTCTTAGTGGACTCGAGTGAGCTGCAGCTGGCCGTG 1236

Qy    310  GluSerProLeuLeuGluGlnValAlaArgArgGlyLysAsnAspAspLeuLeuPheSerIleAla 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1237  GAGTCTCCTCTCCTACTGGAGGTGCGGCGGGGGAAGAACGACGACCTCCTCTTCAGCATCGCA 1296

Qy    330  ProGluValValMet 334
           |||||||||||||||
Db   1297  CCTGAGGTGGTCATG 1311
```

RESULT 5

US-10-275-505-28
; Sequence 28, Application US/10275505
; Publication No. US20040081961A1
; GENERAL INFORMATION:
  APPLICANT: INCYTE GENOMICS, INC.
  APPLICANT: DELEGEANE, Angelo M.; LAL, Preeti G.
  APPLICANT: HAFALIA, April J.A.; PATTERSON, Chandra
  APPLICANT: WALIA, Narinder K.; KEARNEY, Liam
  APPLICANT: TRIBOULEY, Catherine M.; KHAN, Farrah A.
  APPLICANT: YAO, Monique G.; BAUGHN, Mariah A.
  APPLICANT: AZIMZAI, Yalda; ELLIOTT, Vicki S.
  APPLICANT: NGUYEN, Danniel B.; GANDHI, Ameena R.
  APPLICANT: YANG, Junming; HERNANDEZ, Roberto
  APPLICANT: POLICKY, Jennifer L.; LU, Dyung Aina M.
  APPLICANT: REDDY, Roopa M.; YUE, Henry
  APPLICANT: TANG, Y. Tom
  TITLE OF INVENTION: PROTEASES
  FILE REFERENCE: PI-0085 USN
  CURRENT APPLICATION NUMBER: US/10/275,505
  CURRENT FILING DATE: 2002-11-04
  PRIOR APPLICATION NUMBER: PCT/US01/14651
  PRIOR FILING DATE: 2001-05-04
  PRIOR APPLICATION NUMBER: 60/209,402
  PRIOR FILING DATE: 2000-06-01
  PRIOR APPLICATION NUMBER: 60/207,477
  PRIOR FILING DATE: 2000-05-25
  PRIOR APPLICATION NUMBER: 60/205,803
  PRIOR FILING DATE: 2000-05-17
  PRIOR APPLICATION NUMBER: 60/205,566
  PRIOR FILING DATE: 2000-05-11
  PRIOR APPLICATION NUMBER: 60/203,082
  PRIOR FILING DATE: 2000-05-04
  NUMBER OF SEQ ID NOS: 28
  SOFTWARE: PERL Program
  SEQ ID NO 28
  LENGTH: 1797
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  OTHER INFORMATION: Incyte ID No: 7474343CB1
US-10-275-505-28
; Query Match: 97.38%

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

```
1.196=175                Length:        1797
1635.00                  Matches:        325
100.00%                  Conservative:     0
100.00%                  Mismatches:       0
97.38%                   Indels:           0
18                       Gaps:             0
```

US-10-617-443B-2 (1-334) x US-10-275-505-28 (1-1797)

```
Qy     10  GlyLeuHisGlnLeuSerSerProArgTyrLeuPheAsnPheIleAlaAspValValGlu 29
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    565  GGTCTCCACCAGCTGAGCAGCCCCCGCTACCTGTTCAACTTCATTGCTGACGTGGTGGAG 624

Qy     30  LysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625  AAGATCGCACCAGCTGTGGCTCACATAGAGCTCTTCCTGAGACACCCCCTGTTTGGCCGC 684

Qy     50  AsnValProLeuSerSerGlyPheIleMetSerGlyGluValAlaGlyLeuIleIleThr 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    685  AACGTGCCCCTGTCCAGCGGGTTCATCATGTCAGGTGAAGTTGCAGGCCTGATCATCACC 744

Qy     70  AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745  AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGCAGGCAGCAGCTCAAGGTGCAG 804

Qy     90  LeuGlnAsnGlyValAspSerTyrGluAlaThrIleLysAspAlaIleAspSerLysAsp 109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805  CTACAGAATGGGGTGGACACTTCCTATGAGGCCACCATCAAAGACGCCATCGACAGCAAA 864

Qy    110  AlaThrIleLysIleIleHisProLeuValLysLysLeuProValAlaHisSerAla 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865  GCCACCATCAAGATCATCCACCCACTCGTGAAAAAGCTCCCCGTGTTGCGTGGGTCTGCC 924

Qy    130  AspLeuArgProGlyIleGluPheValAlaIleAlaLeuGlySerProPheAlaLeuGlnLeuAsnThr 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925  GACCTGCGGCCGGGCATTGAGTTTGTGGCCATCGCCCTACACAGAACACA 984

Qy    150  ValThrThrGlyIleLeuValSerThrArgAlaGlnArgGluGlyLeuGlyLeuArgAsp 169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985  GTGACAACGGGCATCGTCAGCACTCGCCCAGCGGAGGGCCTCGGGCTCCGGGAC 1044

Qy    170  SerAspMetAspTyrIleGlnThrAspAlaIleIleAlaAsnTyrGlyAsnSerGlyGlyPro 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1045  TCCGACATGGACTACATCCAGACGGATGCCATCATCATCAACTACGGGAACTCCGGGGGACCA 1104

Qy    190  LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105  CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC 1164

Qy    210  SerPheAlaIleProSerAspAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165  TCCTTTGCCATCCCCTCAGACGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 1224

Qy    230  IleLeuGlyAsnTrpLysSerArgPheIleGlyIleIleArgMetArgThrIleThrProSerLeu 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225  ATCCAAAGACTGGAAAGCAGGAGGCCTTCATCGGCATACCGGATGCGGACCATCACACCAAGCCTG 1284

Qy    250  ValAspSerLysLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1285  GTGGATAGCAGCTGAAGGCCCAGCAACCCGGACTTCCCAGAGGTCAGCAGTGGGAATTTATGTG 1344

Qy    270  GlnGluValAlaProAsnSerProGlnArgGlyIleGlnAspGlyIleIleTyrIle 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345  CAAGAGGTTGCGCCGAATTCACCTTCCACGAGGAGGCCATCCAAGATGGGTGACATCATC 1404

Qy    290  ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnLeuAlaValLeuThr 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1405  GTCAAGGTCAACGGGCGGCCGGTCCTCTTAGTGGACTCGAGTGAGCTGCAGCTGGCCGTG 1464

Qy    310  GluSerProLeuLeuGluGlnValAlaArgArgGlyLysAsnAspAspLeuLeuPheSerIleAla 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1465  GAGTCTCCTCTCCTACTGGAGGTGCGGCGGGGGAAGAACGACGACCTCCTCTTCAGCATCGCA 1524

Qy    330  ProGluValValMet 334
           |||||||||||||||
Db   1525  CCTGAGGTGGTCATG 1539
```

RESULT 6
US-09-818-143-14
; Sequence 14, Application US/09818143

```
; Patent No. US20020019000A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES
; FILE REFERENCE: PB-0004 CIP
; CURRENT APPLICATION NUMBER: US/09/818,143
; CURRENT FILING DATE: 2001-03-26
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2814981CB1
US-09-818-143-14

Qy Match:            9

Alignment Scores:
Pred. No.:           1.43e-175
Score:               1635.00
Percent Similarity:  100.00%
Best Local Similarity: 97.38%
Query Match:         9

                     Length:        2040
                     Matches:       325
                     Conservative:  0
                     Mismatches:    0
                     Indels:        0
                     Gaps:          0

US-10-617-443B-2 (1-334) x US-09-818-143-14 (1-2040)
```

```
Qy    10 GlyIleuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72 GGTCTCCACCAGTTGAGCAGCCCCGCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 131

Qy    30 LysIleAlaProAlaValAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 191

Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyTyrIleIleThr 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 AACGTGCCCCTGTCCCAGCGGGTTCTGGCTTCATCATGTCAGAGGCCGGCTGACATCACC 251

Qy    70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   252 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGAGGCCAGCTCAAGGTGCAG 311

Qy    90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleThrLysAspLysLysSerAspIle 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   312 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAAGACAATCGAAGAGTCGGACATT 371

Qy   110 AlaThrIleLysIleHisProLysGlyLysLeuProValLeuLeuGlyHisSerAla 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372 GCCACCATCAAAGATCCACCAAGAAAAAAGCTCCCTGTGTTGTTGCTGGTCACTCGGGCC 431

Qy   130 AspLeuArgProGlyPheValAlaAlaIleGlySerProPheAlaLeuGluGlnAsnThr 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   432 GACCTGCGCCGGCCTGGGGAGTTTGTGCGGGCCCATCGGCCATCCCTTCGCCCTACAGAACACA 491

Qy   150 ValThrThrGlyIleValSerThrAlaGlnAsnArgGluLysGlyLeuArgAsp 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   492 GTGACAACCGGCATCGTCAGCACTGCCCAGGAGAACCTGGGCCCTCCGGAC 551

Qy   170 SerAspMetAspTyrIleGlnThrAspAlaIleIleIleAlaAsnTyrGlyAsnSerGlyGlyPro 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   552 TCCGACATGGACTACATCCAGACACGGGAAGCTCATCATCAACTACGGGAACTCCGGGGGACCA 611

Qy   190 LeuValAlaAsnLeuAspProGlyGlyValIleGlyIleIleAsnThrLeuThrAlaGlyTyrIle 209
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   612 CTGGTGGCCAACCTGGATCCGGGAGGTCATTGGCCATCAACACGGTCACGGCTGGCTATATC 671

Qy   210 SerPheAlaIleIleProSerAspArgIleThrArgGluPheGlnAspLysGln 229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   672 TCCTTTGCCATCCCCTCAGACCGCATCACACGGGTTCCTACAGAGTTCCAAGACAAGCAG 731

Qy   330 ProGluValValMet 334
         ||||||||||||||
Db  1032 CCTGAGGTGGTCATG 1046
```

```
Qy   230 IleLysAspTrpLysLysArgPheIleGlyIleuLeuArgMetArgThrLeuThrProSerLeu 249
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   732 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATACGGATGCGGACGACACACCAAGCCTG 791

Qy   250 ValAspGluLeuLysLysAlaSerAsnProAlaAspPhePheProGluValSerSerGlyIleTyrVal 269
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   792 GTGGATGAGCTGAAGAAGGCCAGCAACCCCGACTTCCCAGAGGTCAGCTGGAATTTATGTG 851

Qy   270 GlnLeuValAlaLaProAsnSerProSerGlnArgGlyIleIleGlnAspIleIleIle 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   852 CAAGAGGTTGCGCGGCCGGAATTCACCTTCTCAGAAGACCGGGATCGAGGATGACATCATC 911

Qy   290 ValLysAspAsnGlyLysArgProLeuValAlaSerProLeuLeuGlnLeuuValAlaValLeuThr 309
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   912 GTCAAGGTGCAACGGCGGTCCTCTAGTGGACTCGAGTCGCAGGAGGGCGTGCTCAAGGTGACT 971

Qy   310 GluSerProLeuLeuLeuGluValArgGlyValArgGlyValAsnAspAspLeuLeuPheSerIleAla 329
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   972 GAGTCTCCTCTCCTTCTAGTCGGAGGTGCGGGCGGGGGAAACGACGACCTCCTCTTCAGCATCGCA 1031
```

```
; RESULT 7
; US-10-381-820A-7
; Sequence 7, Application US/10381820A
; Publication No. US20040142333A1
; GENERAL INFORMATION:
; APPLICANT: Deshun, Lu
; APPLICANT: Song, Ho Yeong
; APPLICANT: Su, Eric Wen
; APPLICANT: Wang, He
; TITLE OF INVENTION: Novel Secreted Proteins and Their Uses
; FILE REFERENCE: X-13974
; CURRENT APPLICATION NUMBER: US/10/381,820A
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7
; LENGTH: 2094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (206)..(1564)
US-10-381-820A-7

Alignment Scores:
Pred. No.:           1.499e-175
Score:               1635.00
Percent Similarity:  100.00%
Best Local Similarity: 97.38%
Query Match:         19

                     Length:        2094
                     Matches:       325
                     Conservative:  0
                     Mismatches:    0
                     Indels:        0
                     Gaps:          0

US-10-617-443B-2 (1-334) x US-10-381-820A-7 (1-2094)
```

```
Qy    10 GlyIleuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   590 GGTCTCCACCAGTTGAGCAGCCCCGCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 649

Qy    30 LysIleAlaProAlaValAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   650 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 709

Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyTyrIleIleThr 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   710 AACGTGCCCCTGTCCCAGCGGGTTCTGGCTTCATCATGTCAGAGGCCGGCTGACATCACC 769

Qy    70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   770 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGAGGCCAGCTCAAGGTGCAG 829
```

```
Qy   90   LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspIleLysLysSerAspIle 109
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  830   CTACAGAATGGGAGCTCCTATGAGGCCACCATCAAAGACATCGACAGAAGTCGGACATT 889

Qy  110   AlaThrIleLysHisIleProLysLysSerLeuLeuLeuLeuGlyHisSerAla 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  890   GCCACCATCAAGATCCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTGTCACTCGGCC 949

Qy  130   AspLeuArgProGlyGluProPheAlaLeuGlnAsnThr 149
          |||||||||||||||||||||||||||||||||||||||
Db  950   GACCTCCGGCCTGGGGAGTTTGTTGCTGGCCATCGGAGTCGGACACA 1009

Qy  150   ValThrThrGlyIleLeuValSerThrAlaGlnAsnGluGlyLeuGlyLeuArgAsp 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1010   GTGACAACGGGCATCCTCGTCAGCACTGCCCAGAACGGGAGGGAGCTGGGGCTCCGGGAC 1069

Qy  170   SerAspMetAspThrTyrIleGlnThrAspAlaIleIleAsnThrTyrGlyGlyPro 189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1070   TCCGACATGGACACTTACATCCAGACAGACGCCATCATCAACACTTACGGGGGCC 1129

Qy  190   LeuValAsnLeuAspGlyGluValAlaIleGlyIleIleGluAsnThrLeuLysValThrAlaGlyIleIle 209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1130   CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC 1189

Qy  210   SerPheAlaIleProSerAspAspArgIleThrArgPheLeuThrGluLysAspSerGln 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1190   TCCTTTGCCATCCCCTCAGACGGCCATCAACAGGTTCCTCACAGAGTTCCAAGAACAAGCAG 1249

Qy  230   IleLeuLysThrLysLysArgPheIleGluIleTyrMetArgThrIleThrProSerLeu 249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1250   ATCAAAGACTGGAAGAAGCGCTTCATCGGGCATACGGAACACCAACCAAGCCTG 1309

Qy  250   ValAspGlyGluLysLysAlaSerArgAsnProAspPheProGluValSerSerGlyIleTyrVal 269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1310   GTGGATGAGCTGAAGGCCCAGACCAACCCGGACTTCCCAGAGGTCAGCAGTGGGATTTATGTG 1369

Qy  270   GlnGluValAlaProAlaAsnSerProSerGlnAlaArgGlyGlyTyrIleGluIleAsnGlyTyrAspIleIle 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1370   CAAGAGGTTGCGCCCGCAAATTCACCTTCACAGGCGCGGGGCATCGAGATAAAGGTGGTGACATCATC 1429

Qy  290   ValLysSerValAlasnGlyArgProLeuValAspSerSerGluLeuGlnAlaValLeuThr 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1430   GTCAAGGTCAACGGGCGGCCTCTCGAGTGACTGCAGGAGCCCGTGCTTGACC 1489

Qy  310   GluSerProLeuLeuGluLeuGluValArgAsnAsphapLeuLeuPheSerIleAla 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1490   GAGTCTCCTCTCCTACTGGAGCTGGAGGTGCCGGGGGAACGACGACCTCCTCTTCAGCATCGCA 1549

Qy  330   ProGluValValMet 334
          |||||||||||||||
Db 1550   CCTGAGGTGGTCATG 1564
```

RESULT 8

US-10-956-157-5181

Sequence 5181, Application US/1095615 7

Publication No. US20050116625A1

```
Alignment Scores:
Pred. No.:                        1.97e-175      Length:          2541
Score:                            1635.00        Matches:          325
Percent Similarity:               100.00%        Conservative:       0
Best Local Similarity:            100.00%        Mismatches:         0
Query Match:                       97.38%        Indels:             0
DB:                                    21        Gaps:               0

US-10-617-443b-2 (1-334) x US-10-956-157-5181 (1-2541)

Qy   10   GlyLeuHisGlnLeuSerSerProArgTyrlysPheAsnPheIleAlaAspValValGlu 29
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  589   GGTCTCCACCAGCTGAGCAGCCCGCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 648

Qy   30   LysIleAlaProAlaValValHisIleGluLeuProheLeuArgHisProLeuPheGlyArg 49
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  649   AAGATCGCACCAGCCGTTGTCCACATAGAGCTTCCTTGAGCACCCCCTGTTTGGGCGC 708

Qy   50   AsnValProLeuSerGlySerGlyTyrGlyPheIleMetSerGluIleAlaGlyLeuIleThr 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  709   AACGTGCCCCTGTCCAGCGGGTTCTGGCTTCATCATGTCAGAGATCATCACC 768

Qy   70   AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnLeuLysValGln 89
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  769   AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGCCGCCAGCTCAAGGTGCAG 828

Qy   90   LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspIleLysLysSerAspIle 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  829   CTACAGAATGGGAGCTCCTATGAGGCCACCATCAAAGACATCGACAGAAGTCGGACATT 888

Qy  110   AlaThrIleLysHisIleProLysLysSerLeuLeuLeuLeuGlyHisSerAla 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  889   GCCACCATCAAGATCCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCC 948

Qy  130   AspLeuArgProGlyGluProPheAlaLeuGlnAsnThr 149
          |||||||||||||||||||||||||||||||||||||||
Db  949   GACCTCGGACTACAATCCCAGACAGTGGGCACAGAACAGACACA 1008

Qy  150   ValThrThrGlyIleLeuValSerThrAlaGlnAsnGluGlyLeuGlyLeuArgAsp 169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1009   GTGACAACGGGCATCCTCGTCAGCACTGCCCAGAACGGGAGGGAGCTGGGGCTCCGGGAC 1068

Qy  170   SerAspMetAspThrTyrIleGlnThrAspAlaIleIleAsnThrTyrGlyGlyPro 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1069   TCCGACATGGACACTTACATCCAGACAGACGCCATCATCAACACTTACGGGGGCCA 1128

Qy  190   LeuValAsnLeuAspGlyGluValAlaIleGlyIleIleGluAsnThrLeuLysValThrAlaGlyIleIle 209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1129   CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC 1188

Qy  210   SerPheAlaIleProSerAspAspArgIleThrArgPheLeuThrGluLysAspSerGln 229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1189   TCCTTTGCCATCCCCTCAGACGACCGCATCAACAGGTTCCTCACAGAGTTCCAAGAACAAGCAG 1248

Qy  230   IleLeuLysThrLysLysArgPheIleGluIleTyrMetArgThrIleThrProSerLeu 249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1249   ATCAAAGACTGGAAGAAGCGCTTCATCGGGCATACGGATACCAACCAAGCCTG 1308

Qy  250   ValAspGlyGluLysLysAlaSerArgAsnProAspPheProGluValSerSerGlyIleTyrVal 269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1309   GTGGATGAGCTGAAGGCCCAGACCAACCCGGACTTCCCAGAGGTCAGCAGTGGGAATTTATGTG 1368

Qy  270   GlnGluValAlaProAlaAsnSerProSerGlnAlaArgGlyGlyTyrIleGluIleAsnGlyTyrAspIleIle 289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1369   CAAGAGGTTGCGCCCGCAAATTCACCTTCACAGGCGCGGGGCATCGAGATAAAGGTGGTGACATCATC 1428

Qy  290   ValLysSerValAlaSnGlyArgProLeuValAspSerSerGluLeuGlnAlaValLeuThr 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1429   GTCAAGGTCAACGGGCGGCCTCTCGAGTGACTGCAGGAGCCCGTGCTTGACC 1488

Qy  310   GluSerProLeuLeuGluLeuGluValArgAsnAspAspLeuLeuPheSerIleAla 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1489   GAGTCTCCTCTCCTACTGGAGCTGGAGGTGCCGGGGGAACGACGACCTCCTCTTCAGCATCGCA 1548
```

```
Qy   330 ProGluValValMet 334
         |||||||||||||||
Db  1549 CCTGAGGTGGTCATG 1563
```

RESULT 9
US-10-485-313A-31
; Sequence 31, Application US/10485313A
; Publication No. US2005005902A1
; GENERAL INFORMATION:
; APPLICANT: NIE, Guiying
; APPLICANT: SALAMONSEN, Lois Adrienne
; APPLICANT: LI, Ying
; APPLICANT: HAMPTON, Anne Lorraine
; APPLICANT: FINDLAY, John Kerr
; TITLE OF INVENTION: Novel Serine Protease
; FILE REFERENCE: 31633-200357
; CURRENT APPLICATION NUMBER: US/10/485,313A
; CURRENT FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: PCT/AU02/01010
; PRIOR FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: PR6707
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 31
; LENGTH: 2543
; TYPE: DNA
; ORGANISM: Homo sapiens
; DB:

US-10-617-443B-2 (1-334) x US-10-485-313A-31 (1-2543)

Alignment Scores:
Pred. No.:
Score: 1.97e-175
                       1635.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 97.38%
DB: 21

Length: 2543
Matches: 325
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0
```

```
Qy    10 GlyLeuHisGlnLeuSerSerProArgThrTyrLysPheAsnPheIleAlaAspValValGlu 29
Db   577 GGTCTCCACCAGTGAGCAGCCCCGCGTTACAAGTTCAATTGCTGACGTGGTGGAG 636

Qy    30 LysIleLeuAProAlaValValHisIleGlyLeuPheLeuArgHisProLeuPheGlyArg 49
Db   637 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTCAGACACCCCGCTGTTTGGCCGC 696

Qy    50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
Db   697 AACGTGCCCCTGTCCAGCGGGTTCATCATGTCAGAGGCCGGCCTGATCATCACC 756

Qy    70 AsnAlaHisValValSerSerAsnSerArgLeuAlaLaaProGlyArgGlnGlnLeuValValGln 89
Db   757 AATGCCCACGTGGTGTCCAGCAACAGTGCTCGCCCGGGAGGCTCAAGGTGCAG 816

Qy    90 LeuGlnAsnGlyAspSerTyrGluValaThrThrIleLeuAspIleAspLysLysSerAspIle 109
Db   817 CTACAGAATGGGGACTCCTATGAGGCCACCACATCAAAGACATCGACAAGAAGAGTGGACATT 876

Qy   110 AlaThrIleLysIleHisProLysGlySerProValLeuLeuLeuGlyHisSerAla 129
Db   877 GCCACCATCAAGATCCATCCCAAGAAAAAAGCTCCCGTGTTGTTGTTGGCTGGGTCACTCGCC 936

Qy   130 AspLeuArgProGlyPheValValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
Db   937 GACCTGCGGCCTGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 996

Qy   150 ValThrThrGlyIleValSerThrAlaGlnArgGluGlyLeuArgAsp 169
Db   997 GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGGCTGCGGGAC 1056
```

```
Qy   170 SerAspMetAspTyrIleGlnThrAspAlaIleIleIleAsnTyrGlyAsnSerGlyGlyPro 189
Db  1057 TCCGACATGGACTACATCCAGACCGATGCCATCATCATCAACTACGGGAACTCCGGGGGACCA 1116

Qy   190 LeuValAsnLeuAspGlyGluValIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
Db  1117 CTGGTGAACCTGGATGGGGAGGTCATTGGCATCAACACGCTCAAGGTGACAGCTGGCATC 1176

Qy   210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnGln 229
Db  1177 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAGCAGCAG 1236

Qy   230 IleLeuLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
Db  1237 ATCAAAGACTGGAAGAAGCGCTTCATCGGCATCCGCATGCGCACCATCACACCAAGCCTG 1296

Qy   250 ValAspGlyLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
Db  1297 GTGGATGGGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTGTCCAGTGGAATTTATGTG 1356

Qy   270 GlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnAsnGlyAspIleIle 289
Db  1357 CAAGAGGTGGCCCCGAATTCACCTTCTCAGAGAGGCGGCCATCCAGAATGGTGACATCATC 1416

Qy   290 ValLysValAsnGlyArgProLeuValAspSerSerGluLeuGlnLnGluAlaValLeuThr 309
Db  1417 GTCAAGGTCAACGGGCGTCCTCTAGTGGACTCGAGTGAGCTGCAGGAGGCCGTGCTGACC 1476

Qy   310 GluSerProLeuLeuGluLysLeuArgProGlyArgGlyAsnAspAspSerLeuPheSerIleAla 329
Db  1477 GAGTCTCCTCTCCTACTGGAGAAGCTGCGGCCGGGAAGCGACGACTCCTCTTCAGCATCGCA 1536

Qy   330 ProGluValValMet 334
         |||||||||||||||
Db  1537 CCTGAGGTGGTCATG 1551
```

RESULT 10
US-09-796-753-31
; Sequence 31, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: US/09/796,753
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-09-30

```
;  PRIOR FILING DATE: 1999-12-23
;  PRIOR APPLICATION NUMBER: 09/474,071
;  PRIOR FILING DATE: 1999-12-29
;  PRIOR APPLICATION NUMBER: 09/474,072
;  PRIOR FILING DATE: 1999-12-29
;  PRIOR APPLICATION NUMBER: 09/514,010
;  PRIOR FILING DATE: 2000-02-25
;  PRIOR APPLICATION NUMBER: 09/516,745
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 09/572,002
;  PRIOR FILING DATE: 2000-05-14
;  PRIOR APPLICATION NUMBER: 09/597,993
;  PRIOR FILING DATE: 2000-06-19
;  PRIOR APPLICATION NUMBER: 09/599,596
;  PRIOR FILING DATE: 2000-06-22
;  PRIOR APPLICATION NUMBER: 09/630,334
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 09/606,565
;  PRIOR FILING DATE: 2000-06-29
;  PRIOR APPLICATION NUMBER: 09/606,317
;  PRIOR FILING DATE: 2000-06-29
;  PRIOR APPLICATION NUMBER: 09/665,666
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: 09/677,751
;  PRIOR FILING DATE: 2000-09-30
;  NUMBER OF SEQ ID NOS: 162
;  SEQ ID NO 31
;  LENGTH: 2576
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-796-753-31

Alignment Scores:
Pred. No.:                    2.01e-175
Score:                        1635.00
Percent Similarity:           100.00%
Best Local Similarity:        100.00%
Query Match:                  97.38%
DB:                           10
```

```
                              Length:          2576
                              Matches:          325
                              Conservative:       0
                              Mismatches:         0
                              Indels:             0
                              Gaps:               0
```

```
US-10-617-443B-2 (1-334) x US-09-796-753-31 (1-2576)

Qy      10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     606 GGTCTCCACCAGCTGAGCAGCCCGCCCTACAAGTTCAACTTCATTGCTGACGTGGTGGAG 665

Qy      30 LysIleAlaProAlaValValHisIleGlnLeuPheLeuArgHisProLeuPheGlyArg 49
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     666 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGC 725

Qy      50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleLeuThr 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     726 AACGTGCCCCTGTCCTCCGGCGGTTCCATCATGTCAGACGCAGGTTTGATCCTGACC 785

Qy      70 AsnAlaHisValSerArgAsnSerAlaAlaProGlyArgGlnLeuValGln 89
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 AATGCCCACGTGGTCTCCAGCAACAGTGCTGCCCCGGGCAGCTAAGGTGCAG 845

Qy      90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleLysSerAspIle 109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCAAGAGTCGGACATT 905

Qy     110 AlaThrIleLysIleHisProLysGlyGlnValLeuLeuLeuGlyHisSerAla 129
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     906 GCCACCATCAAGATCCATCCCAAGAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCC 965

Qy     130 AspLeuArgProGlyPheValAlaAlaIleGlySerProPheAlaLeuGlnAsnThr 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 GACCTGGCGGCCTGGGGAGTTTGTGGCCATCGGCAGTCCCTTACAGAACACA 1025

Qy     150 ValThrThrGlyIleValSerThrAlaGlyAsnArgGluGlyArgGluLeuGlyLeuArgAsp 169
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 GTGACAACCGGCATCGTCAGCACTGCCCAGGGGGAGGGAGCTGGGCCTCGGGGAC 1085
```

```
Qy     170 SerAspMetAspThrIleGlnThrAspAlaIleIleAsnThrGlyAsnSerGlyGlyPro 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 TCCGACATGGACTACCACCCAGAGACGGAGGCCACCATCAACATCACGGGAAACTCCGGGGGGACCA 1145

Qy     190 LeuValAsnLeuAspGlyGluValIleGlyValAsnThrLeuLysValThrAlaGlyIle 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 CTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATC 1205

Qy     210 SerPheAlaIleProSerArgAspArgIleThrArgPheLeuThrGluPheGlnAspLysLysGln 229
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1206 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGAAG 1265

Qy     230 IleLeuLysAspTrpLysLysArgPheIleGlyIleArgMetArgThrIleThrProSerLeu 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1266 ATCAAGGACTGGAAGAAGCGCTTCATCGGCATACGGATGCGGACGATCACACCAAGCCTG 1325

Qy     250 ValAspGluLeuLysAlaSerAsnProAspPheProGluValSerSerGlyIleTyrVal 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1326 GTGGATGAGCTGAAGGCCAGCAACCCGGACTTCCCAGAGGTGTCCAGCTGGGAATTTATGTG 1385

Qy     270 GlnGluValAlaAlaProAsnSerProSerGlnAsnGlyGlyIleGlnAsnGlyAspIleIle 289
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1386 CAAGAGGTTGCCGCGAATTCACCTTCTCAGAGACAGGCGGCCATCCAAGATGGTGACATCATC 1445

Qy     290 ValLysValAsnGlyArgProLeuValAspSerSerGluGlnGlnAlaValValLeuThr 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1446 GTCAAGGTCAACGGGCGGGTCCTCGTTGGAGTGGACTCGAGTGAGCAGCGTGCTGGTGACC 1505

Qy     310 GluSerProLeuLeuGluValArgArgGlyAsnAspAspLeuLeuPheSerIleAla 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1506 GAGTCTCCTCCTCCTCCTACTGGAGGTGCGGCGGGGGGAACCAGCAGACCTTCCTCTTCAGCATCGCA 1565

Qy     330 ProGluValValMet 334
           |||||||||||||
Db    1566 CCTGAGGTGGTCATG 1580
```

```
RESULT 11
US-10-485-313A-26
; Sequence 26, Application US/10485313A
; Publication No. US20050059002A1
; GENERAL INFORMATION:
;  APPLICANT: NIE, Guiying
;  APPLICANT: SALAMONSEN, Lois Adrienne
;  APPLICANT: LI, Ying
;  APPLICANT: HAMPTON, Anne Lorraine
;  APPLICANT: FINDLAY, John Kerr
;  TITLE OF INVENTION: Novel Serine Protease
;  FILE REFERENCE: 31633-200357
;  CURRENT APPLICATION NUMBER: US/10/485,313A
;  CURRENT FILING DATE: 2004-01-30
;  PRIOR APPLICATION NUMBER: PCT/AU02/01010
;  PRIOR FILING DATE: 2002-07-30
;  PRIOR APPLICATION NUMBER: PR6707
;  PRIOR FILING DATE: 2001-07-30
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: Patentin version 3.1
;  SEQ ID NO 26
;  LENGTH: 2450
;  TYPE: DNA
;  ORGANISM: Mus musculus
US-10-485-313A-26

Alignment Scores:
Pred. No.:                    3.56e-164
Score:                        1536.00
Percent Similarity:           96.63%
Best Local Similarity:        92.64%
Query Match:                  91.48%
DB:                           21
```

```
                              Length:          2450
                              Matches:          302
                              Conservative:      13
                              Mismatches:        11
                              Indels:             0
                              Gaps:               0
```

```
US-10-617-443B-2 (1-334) x US-10-485-313A-26 (1-2450)
```

Alignment Scores:
Pred. No.:

Score:	1482.50	Length: 2551
Percent Similarity:	97.55%	Matches: 319
Best Local Similarity:	88.30%	Mismatches: 6
Query Match:	15	Conservative: 0
		Indels: 7
		Gaps: 0

4.71e-158

DB:

US-10-617-443B-2 (1-334) x US-10-301-822-78 (1-2551)

```
Qy    9 AlaGlyLeuHisGlnLeuSerSerProArgThrTyrLysPheAsnPheIleAlaAspValVal  28
          : ::|||||||||||||||||||||||||||||||||||||||||
Db  526 TCTGGCTTCCACCAGCTGAGCAGCCCCGGTACAAGTTCAACTTCATTGCCGATGTGGTG  585

Qy   29 GluLysIleAlaProAlaLysValHisIleIleGluLeuLeuLeuArgHisProLeuPheGly  48
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  586 GAGAAGATTGCGCCAGCTGGTGGTCCACATAGAGCTCTTTCTGAGACACCCCCTGTTTGGC  645

Qy   49 ArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGluValAlaGlyLeuIleIle  68
          :::||||||||||||||||||||||||||||||||
Db  646 CGGAATGTGCCGCTGTCCAGTGGCTTCATCATGTCAGAAGTAGCCGGTTTGATCGTC  705

Qy   69 ThrAsnAlaHisValValSerSerSerAsnSerAlaAlaProGlyArgGlnLeuLeuLysVal  88
          |||||||||||||||||||||||||||||||||||||||
Db  706 ACCAACGCCCACGTGGTCTCCAGCTCCAACTCCGGCCGGCAGCTGCTGAAGGTG  765

Qy   89 GlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysGlySerAsp 108
          ||||||||||||||||||||||||||||||||||||||||
Db  766 CAGCTGCAGAATGGGGACTCCTATGAGGCCACCATCAAGGACATCGACAAGGGGTCGGAC  825

Qy  109 IleAlaThrIleLysIleHisProLysGlyLysGlyLeuProValLeuLeuLeuGlyHisSer 128
          |||||||||||||||||||||||||||||||||||||||
Db  826 ATTGCCACGATTAAATCACCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCA  885

Qy  129 AlaAspSerLeuArgProGlyGlyPheValValAlaIleGluGlySerProPheAlaLeuAsn 148
          ||||||||||||||||||||||||||||||||||||||
Db  886 GCAGACCTGCGCCCTGGCGGAGTTGTGGTTGCCATCGGGAGCCCCTTTGCCCTGAAC  945

Qy  149 ThrValThrThrGlyIleValSerThrAlaGlnGlnArgGluGlyLeuGlyLeuLeuArg 168
          ||||||||||||||||||||||||||||||||||||
Db  946 ACCGTGACAACCGGCCATTGTCAGCACTGCCCAGCAGCGGGAGGGACTGGGTCTCCGG 1005

Qy  169 AspSerAspMetAspThrTyrIleGluThrArgAspIleAsnThrTyrGlyAsnSerGlyGly 188
          |||||||||||||||||||||||||||||||||
Db 1006 GACTCAGACATGGACTACATATACGAGACCCGAGATATCAATACGTACGGGAACTCAGGAGGA 1065

Qy  189 ProLeuValAsnLeuAsnAspGlyIleGluGlyIleLeuAsnLysLysValValThrAlaGly 208
          ||::||||||||||||||||||||||||||||
Db 1066 CCCCTGGTGAACCTGAACGATGGCATCGAGGGTCATCCTGAACAACACGCTCAAGGTTGCAGGC 1125

Qy  209 IleSerSerPheAlaIleIleProSerArgAspArgIleThrArgPheLeuThrGluGlnLys 228
          |||||||||||||||||||||||||||||||||
Db 1126 ATCTCCTTTGCCATCCCTTCACGGGATCGCATCACACGCTTCCTCACTGAGCAGAAAAG 1185

Qy  229 GlnIleLeuLysLysTrpPheLysSerArgPheLeuIleAlaArgMetArgThrIleThrPro 248
          :::||||||||||||||||||||||||||||||
Db 1186 CAGATCTTGAAGAAGACTGGAAGAGCGCTTCAATCCGACCATCACGCCAAGT 1245

Qy  249 LeuValAspGluGluLeuLysValAlaSerAsnProAspAspPheGlnValSerSerGlyTyr 268
          ||||:|||||||||||||||||||||||||||
Db 1246 TTGGTGGACGAACTGAAGCCCGCCAACACTTTCCACGGTGCAGTGGAATATAT 1305

Qy  269 ValGlnGluValAlaAlaProAsnSerProSerGlnArgGlnAspGlyTyrIleGlnAspIle 288
          ||||||||||||||||||||||||||||||||
Db 1306 GTTCAAGAGGTGGTTGCCAATTCACCTTCACCGTGACACCCAGAGATGGGCACATC 1365

Qy  289 IleLeuValLysValAlaSerGlyProLeuGlnGluValAlaValVal 308
          |||||||||||||||||||||||||||||
Db 1366 ATCCTGGTAAAGGTTCAATGCCGTCCAGAGGAGTTCCAGGAGGCACATT 1425

Qy  309 ThrGluSerProLeuLeuLeuGlnValArgGlyAsnAspAspLeuPheSerIle 328
          |||||||||||||||||||||||||||||||||
Db 1426 AACGAGTCTTCACTCCTGCTGCAGGTGCGCGGCGAGGCCAATGATGATCTCTTCAGCATC 1485

Qy  329 AlaProGluValValMet 334
          |||||||||||||
Db 1486 ATCCCTGAGGTGGTCATG 1503
```

```
Qy   10 GlyLeuHisGlnLeuSerSerProArgThrTyrLysPheAsnPheIleAlaAspValValGlu  29
          |||||||||||||||||||||||||||||||||||||||||||
Db  606 GGTCTCCACCAGCTGAGCAGCCCCGGCGCTACAAGTTCAACTTCATTGCTGATGTGGGAG  665

Qy   30 LysIleAlaProAlaLysValHisIleIleGluLeuPheLeuArgHisProLeuPheGlyArg  49
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  666 AAGATCGCACCAGCGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCCCTGTTTGGCCGC  725

Qy   50 AsnValProLeuSerSerGlyPheIleMetSerGlnValAlaGlyLeuIleIleThr  69
          |||||||||||||||||||||||||||||||||||||||
Db  726 AACGTGCCCCTGTCCAGCGGCTTCATCATGTCCCAGGTGGCTGGGCTCATCATCACC  785

Qy   70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnLeuLeuLysValGln  89
          |||||||||||||||||||||||||||||||||||||||||||
Db  786 AATGCCCACGTGGTGTCCAGCAACAGCGCTGCCCCCGGTCGTCAGCTGCTCAAGGTGCAG  845

Qy   90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysAspIleAspLysGlySerAspIle 109
          ||||||||||||||||||||||||||||||||||||||||||
Db  846 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGGGATCGGACATT  905

Qy  110 AlaThrIleLysIleHisProLysGlyLysGlyLeuProValLeuLeuLeuGlyHisSerAla 129
          |||||||||||||||||||||||||||||||||||||||||||
Db  906 GCCACCATCAAGATCCATCCCAAAGGCAAAGGCCTCCCCGTGCTGTTGCTGGGTCACTCGGCC  965

Qy  130 AspLeuArgProGlyGlyPheValValAlaIleGluGlySerProPheAlaLeuAsnThr 149
          |||||||||||:|||||||||||||||||||:|||||||||||
Db  966 GACCTGCGCCGT-GGGGAGTTTGTGGTGGCCATCGGAGT-CCCTTCGCCCTACAGAACACA 1023

Qy  150 ValThrThrGlyIleValSerThrAlaGlnGlnArgGluGlyLeuGlyLeuLeuArgAspAsp 169
          ||||||:||||||:|||||||||||:|||||||||||||||||||
Db 1024 GTGACAAC-GGCATCGT-AGCACTGCCCAGCG-GAGGCAGGAGGGAGCTGGGCTCCGGGAC 1080
```

```
RESULT 13
US-09-764-898-112
; Sequence 112, Application US/09764898
; Patent No. US20020096673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 112
; LENGTH: 1868
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-898-112

Alignment Scores:
Pred. No.:
Score:                    2.42e-146
                          1378.00
Percent Similarity:       98.57%
Best Local Similarity:    98.57%
Query Match:              82.07%
DB:                       9

Length:        1868
Matches:       275
Conservative:  0
Mismatches:    4
Indels:        0
Gaps:          0

US-10-617-443B-2 (1-334) x US-09-764-898-112 (1-1868)
```

```
Qy    170 SerAspMetAspTyrIle-GlnThrAspAlaIleIle-AsnTyrGlyAsnSerGlyLyP   189
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCCGACATGGACTACATACCAGAAGGGATTGCCATCATCCAACTACGGGAACTCCGGGGGAC 1140

Qy    189 roLeuValAsnLeuAspProGlyGluValIleLeuGlyIleLeuAsnThrLeuLysValThrAlaGlyI 209
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CACTCGGTGAACCTGGAGCCCGGAGGTGATTCTGGGCATCCTCAACACCCTCAAGGTCACGGCTGGCA 1200

Qy    209 leSerPheAlaIleLeuProSerArgAspArgIleThrArgPheLeuThrGluPheGlnAspLysG  229
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TCTCCTTTGCCATCCCCTCACGAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGC    1260

Qy    229 lnIleLysAspTrpLysArgPheIleGlyIleLeuArgMetArgThrIleThrProSerL     249
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGATCAAAGACTGTAAGAGGTTCATCGGCATACGGATGCGGACCATCACACCAAGCC         1320

Qy    249 euValAspGluLeuLysAlaSerAsnProAspProGluProGluValSerSerGlyIleLeuTyrV   269
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGGTGGATGAGCTGAAGGCCAGCAACCCCGACTTCCCAGAGGTCAGTTCAGGGAATTTATG      1380

Qy    269 alGlnGluValAlaProAsnSerProSerGlnIleArgGlyIleLeuAsnAspGlyAspIleI    289
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGCAAGAGGTTGCGCCGAATCACCTTCTCAGAGAGGCGGGGCATCCTGAAC               1440

Qy    289 leValIleLysAlaAsnGlyArgProLeuValAspSerSerSerGluLeuGlnLeuAlaValLeuT  309
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TCGTCAAGGTCAACGGGGTCCTCTAGTGGACTCGAGTGAGCTGCAGAGGCCGTGCTGA        1500

Qy    309 hrGluSerProLeuLeuLeuLeuGluValArgArgGlyYAsnAspAspLeuLeuPheSerIleA   329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCGAGTCTCCTCTCTGCTACTGGAGGTGCGGCGGGGGAACGACGACCTCCTCTTCAGCATCG    1560

Qy    329 laProGluValValMet  334
              |||||||||||||||||
Db   1561 CACCTGAGGTGGTCATG  1577
```

```
RESULT 14
US-09-969-384-7
; Sequence 7, Application US/09969384
; Publication No. US20020192749A1
; GENERAL INFORMATION:
; APPLICANT: Moore, et al.
; TITLE OF INVENTION: Human Gene Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT055P1
; CURRENT APPLICATION NUMBER: US/09/969,384
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: PCT/US01/10542
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/236,384
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/194,118
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-384-7

Alignment Scores:
```

```
Qy     96 TyrGluAlaThrIleLysAspAspIleAspLysLysSerAspIleAlaThrIleLysTyrIleHis    115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    143 TATGAGGCGACCATCAAAGACGACATCGACAAGAAGTCGGACATAGCCACCATCAAGATCCAT      202

Qy    116 ProLysGlyLysTyrLeuProValLeuLeuLeuGlyHisSerAlaAspLeuAsnLeuArgProGlyGlu  135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    203 CCCAAGAAAAAGTACCTCCTTGTTGTTTGTTCCTGGGGTCACTCGGCCGACCTGGGGAG          262

Qy    136 PheValAlaIleAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleVal      155
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    263 TTTGTGGTTGCCATCGCCATCGGCTCCCTTTCGCCCTACAGAACACAGTGACAAGGGGCATCGTC   322

Qy    156 SerThrAlaGlyHisArgGluGlyTyrArgGluLeuGlyLeuArgGlyAspSerAspMetAspTyrIle 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    323 AGCACTGCCCAGCGGGAGGGGGAGCGGGCCCTCGGGGACTCCGACATG                    382

Qy    176 GlnThrAspAlaIleIleIleAspTyrGlyGlyAsnSerGlyIleGlyProLeuValAspGly      195
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    383 CAGACGGATGCCATCATCATCGACTACGGGGGCAACTCCGGGGATCCCCTGGTGGACGGC        442

Qy    196 GluValIleLeuGlyIleLeuAsnThrLeuLysValValThrAlaGlyIleLeuSerProPheAlaLeuProSer 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    443 GAGGTCATTGGCATCAACTGGCTCAAGGTGGGACATCCTCTCTTTGCCATCCCCTCA           502

Qy    216 AspAspArgIleThrArgPheLeuProGluLeuThrGluGlnIleLysLysAspIleSerLys      235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    503 GACGGCATCACACGGTTCCTCACAGATTCCAAGACCAGATCAAAAAGGAC                  562

Qy    236 ArgPheIleGlyIleLeuArgMetArgThrIleThrProSerLeuValAspGluLeuLysAla     255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563 CGCTTCATCGGCATCTTACGGATGCGGACAATCACCCCAAGCCTGGTGGATGAGCTGAAGGCC     622

Qy    256 SerAsnProAspProAspPheProGluValSerSerGlyIleLeuTyrValGlnGluValAlaProAsn 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    623 AGCAACCCGGACTTCCCAGAGGTCAGCAGTGGAATTTATGTCCAAGAGGTTGCGCCGAAT        682

Qy    276 SerProSerGlnAsnArgGlyIleLeuIleAspGlyAspIleIleLeuLysValLysValLysAsnGlyArg 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    683 TCACCTTCTCAGAAGAGGCGGGCATCCAAGATGGTGACATCATCGTGTCAAGGTCAGT         742

Qy    296 ProLeuValAspSerSerGluLeuGlnLeuGlnLeuAlaValLeuThrGluSerProLeuLeuLeu    315
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    743 CCTCTAGTGGACTCGAGTGAGCTGCAGCTGCAGGCCGTGCTCACCGAGTCTCCTCTGCTACTG     802

Qy    316 GluValArgArgGlyYAsnAspAspLeuLeuPheSerIleAlaProGluValValMet         334
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db    803 GAGGTGCGGCGGGGGGAACGACGACCTCCTCTTCAGCATCGCACCTGAGGTGGTCATG          859
```

```
US-10-617-443B-2 (1-334) x US-09-764-898-112 (1-1868)

Qy     56 GlySerGlyPheIleMetSerGlyAlaGlyLeuIleIleIleThrArgAsnAlaHisValValSer    75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 GGTTCTGGCTTCATCATGTCAGGAGCCGGCCTGATCATCACCAATGCCCACGTGGTGTCC        82

Qy     76 SerAsnSerAlaAlaProGlyArgGlnLeuLysValGlnLeuGlnAsnGlyAspSer          95
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     83 AGCAACAGTGCTGCMCCGGGCGCAGSCAGCAGCTCAAGGTGCAGSTACAGAATGGGGACTCC     142

Alignment Scores:
```

Pred. No.:   3.11e-117
Score: 1122.00
Percent Similarity: 80.25%
Best Local Similarity: 76.18%
Query Match: 66.83%
DB:

Length: 1695
Matches: 243
Conservative: 13
Mismatches: 33
Indels: 30
Gaps: 6

US-10-617-443B-2 (1-334) x US-09-969-384-7 (1-1695)

```
Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu  29
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  229 GGTCTCCACCAGCTTGAGCAGCCCCGCTACACAGTTCAACTTCATTGCTGACGTGGTGGAG  288

Qy   30 LysIleAlaProAlaValAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg  49
Db  289 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCCCTGTTTGGCCGC  348

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluValAlaGlyLeuIleIleThr  69
Db  349 AACGTGCCCCTGTCCAGCGGGTTCGGCTTCATCATGTCAGAGGTCGCCGGCCTGATCATCACC  408

Qy   70 AsnAlaHisValValSerSerAsnSerAlaLeuAlaProGlyGlnGlnLeuLysValGln  89
Db  409 AATGCCCACGTGGTGTCCAGCAACAGTGCTCTGGCCCCGGGACAGCTCAAGGTGCAG  468

Qy   90 LeuGlnAsnGlyAspSerTyrGluValGluLeuAspLysLysGlySerAspAspIle  109
Db  469 CTACAGAATGGGGACTCCTATGAGGTCGAGCTGGACAAGAAGGGCAGTGACGACATT  528

Qy  110 AlaThrIleLysIleHisProLysLysSerLeuProValLeuLeuLeuGlyHisSerAla  129
Db  529 GCCACCATCAAGATCCATCCCAAGAAAAGCCTCCCTGTGTTGTTGCTCGGTCACTCGGCC  588

Qy  130 AspLeuArgProGlyValPheValAlaValAlaIleIleGlySerProPheAlaLeuGlnAsnThr  149
Db  589 GACCTCCGGCCCGGGGAGTTTGTTGCTGTGGCAGTCCCTTCGGCCCTACAGAACACA  648

Qy  150 ValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyArgGlnLeuGlyLeuArgAsp  169
Db  649 GTGACAACGGGCATCGTCAGCACTGCCCAGGGCAGGGAGGCTGGGCGTCCTCGGGAC  708

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaAlaIleLeuGluAsnTyrGlyAsnSerGlyVal  189
Db  709 TCCGACATGGACTACATCCAGACAGGGGAACTCCGGGGAGACCA  768

Qy  190 LeuValLeuAsnLeuAspGlyGlyValIleGlyIleLeuAsnThrLeuLysValThrAlaGlyIle  209
Db  769 CTGGTGGAACCTGGATGGCGAGGTCATCGGCATCAACACGCTCAAGGTCACGGCTGGCATC  828

Qy  210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGlyGln  229
Db  829 TCCTTTGCCATCCCCTCAGACAGGATCACCAGAGTTCCTGACAGAGTTCCAAGACAAGCAG  888

Qy  230 IleLeuLysAspAspTrpLysLysArgPheIleGlyLysIleArgMetArgThrIleThrProSerLeu  249
Db  889 ATCAAA----------------------------------GCCCCCTCACTG  906

Qy  250 ---ValAspGluLeuLysSerAlaSerAsnProAspPheGluValSerSerGlyIleTyr  268
Db  907 GCAGTTCATTGACGAGCAGGGGGCTTCCTCACGTTTCCCCGTCCCATGACCCCGTCAGC  966

Qy  269 ValGlnGluValAlaProAsnSerPro----------SerGlnArg  280
Db  967 CAAGCACATAGTGACCCCAGTGCAGCGCTGGTCGTCCATGAGGGGCTGCACATGAGA  1026

Qy  281 GlyGlyLyrIleGlnAspGlyIleLeuGluValLysAsnGlyArgProLeu----ValAsp  299
Db 1027 GCTGCTGTTGTGAGGATGCCGCCAGCCCCATTGCTTCCTTCCTGTGT  1077

Qy  300 SerSerGluLeuGlnLeuGlnAlaValLeuThrGlu--------SerProLeuLeuLeu  315
Db 1078 AATCTGGAGCCAGGGAGAGAGCCTGTGCTTTCCCAAAGAAGCTGAAGTCTTCTTCTCTG  1134
```

Alignment Scores:
Pred. No.:   1.42e-116
Score: 1117.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 66.53%
DB:

Length: 1953
Matches: 222
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-617-443B-2 (1-334) x US-10-485-313A-32 (1-1953)

```
Qy   10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu  29
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492 GGTCTCCACCAGCTTGAGCAGCCCCGCTACACAGTTCAACTTCATTGCTGACGTGGTGGAG  551

Qy   30 LysIleAlaProAlaValAlaValHisIleGluLeuPheLeuArgHisProLeuPheGlyArg  49
Db  552 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCCCTGTTTGGCCGC  611

Qy   50 AsnValProLeuSerSerGlySerGlyPheIleMetSerGluValAlaGlyLeuIleIleThr  69
Db  612 AACGTGCCCCTGTCCAGCGGGTTCGGCTTCATCATGTCAGAGGTCGCCGGCCTGATCATCACC  671

Qy   70 AsnAlaHisValValSerSerAsnSerAlaLeuAlaProGlyGlnGlnLeuLysValGln  89
Db  672 AATGCCCACGTGGTGTCCAGCAACAGTGCTCTGGCCCCGGGACAGCTCAAGGTGCAG  731

Qy   90 LeuGlnAsnGlyAspSerTyrGluValGluLeuAspLysLysGlySerAspAspIle  109
Db  732 CTACAGAATGGGGACTCCTATGAGGTCGAGCTGGACAAGAAGGGCAGTGACGACATT  791

Qy  110 AlaThrIleLysIleHisProLysLysSerLeuProValLeuLeuLeuGlyHisSerAla  129
Db  792 GCCACCATCAAGATCCATCCCAAGAAAAGCCTCCCTGTGTTGTTGCTCGGTCACTCGGCC  851

Qy  130 AspLeuArgProGlyValPheValAlaValAlaIleIleGlySerProPheAlaLeuGlnAsnThr  149
Db  852 GACCTCCGGCCCGGGGAGTTTGTTGCTGTGGCAGTCCCTTCGGCCCTACAGAACACA  911

Qy  150 ValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyArgGlnLeuGlyLeuArgAsp  169
Db  912 GTGACAACGGGCATCGTCAGCACTGCCCAGGGCAGGGAGGCTGGGCGTCCTCGGGAC  971

Qy  170 SerAspMetAspTyrIleGlnThrAspAlaAlaIleLeuGluAsnTyrGlyAsnSerGlyVal  189
Db  972 TCCGACATGGACTACATCCAGACAGGGGCCGCCATCCTGGAAAACTACGGGAACTCCGGGGTCA  1031

Qy  190 LeuValLeuAsnLeuAspGlyGlyValIleGlyIleLeuAsnThrLeuLysValThrAlaGlyIle  209
```

```
Db  1032  CTGGTGAACCTGGATGGCGAGGTCAATTGGCATCAACACGCTCAAGGTCACCGGCTGGCATC  1091
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   210    SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln  229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1092  TCCTTTGCCATCCCTTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG  1151
          ||||||||||||
Qy   230    IleLys  231
          ||||||
Db  1152  ATCAAA  1157
```

Search completed: June 30, 2005, 01:51:01
Job time : 1067.82 secs

THIS PAGE BLANK (USPTO)

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 21:07:15 ; Search time 5771.48 Seconds
(without alignments)
2202.808 Million cell updates/sec

Title: US-10-617-443B-2
Perfect score: 1679
Sequence: 1 MHLALPASAGLHQLSSPRYK.........LEVRRGNDDLLFSIAPEVVM 334

Searched: 34239544 seqs, 19032134700 residues

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10617443/runat 29062005 135932 26076/app_query.fasta_1.718
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10617443 @CGN 1.1.4385 @runat 29062005 135932 26076 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1528	91.0	2730	3	AK035194	Mus muscu
2	1268	75.5	941	6	BQ675502	AGENCOURT
c 3	1256	74.8	810	6	CD635470	AGENCOURT
4	1216.5	72.5	1854	5	CR731516	Tetraodon
5	1191	70.9	890	5	BQ880316	AGENCOURT
6	1176	70.0	724	6	CD635467	AGENCOURT
7	1157	68.9	757	2	CN787405	AGENCOURT
8	1117	66.5	1990	3	CR604057	full-leng
9	1072.5	63.9	796	7	CO404217	AGENCOURT

10	1061	63.2	798	7	CV106197	AGENCOURT
11	1034.5	61.6	973	9	AY406047	Homo sapi
12	1034.5	61.6	1655	3	CR590731	full-leng
13	1034.5	61.6	1911	3	BC011352	Homo sapi
14	1034.5	61.6	1931	3	CR603877	full-leng
15	1034.5	61.6	1951	3	CR624458	full-leng
16	1034.5	61.6	1992	3	CR623030	full-leng
17	1034.5	61.6	2002	3	CR605013	full-leng
18	1034.5	61.6	2005	3	CR593682	full-leng
19	1034.5	61.6	2035	3	CR597732	full-leng
20	1031	61.4	792	7	CV106456	AGENCOURT
21	1030.5	61.6	1753	3	AK090320	Mus muscu
22	1026.5	61.1	973	9	AY406049	Mus muscu
23	1009	60.1	636	7	CN787739	4121948 B
24	990	59.0	852	7	CK599706	AGENCOURT
25	964	57.4	881	6	AK046614	Mus muscu
26	961	57.2	881	6	CD518132	AGENCOURT
27	955.5	56.9	1137	4	BM461040	AGENCOURT
28	953	56.8	745	7	BQ572908	UI-M-FD0-
29	948	56.5	745	7	CO386615	AGENCOURT
30	945	56.3	1328	3	AK008764	Mus muscu
31	943	56.2	616	9	BQ328404	Mus muscu
32	936	55.8	589	7	CV348463	CM3-LiT006
c 33	934	55.7	609	6	CD635468	56056242J
34	933	55.6	557	9	CV348463	MR2-HN003
35	904	53.8	650	1	AY403125	Homo sapi
36	901.5	53.7	1360	3	AU296610	AGENCOURT
37	898	53.5	730	7	CK479697	AGENCOURT
38	892	53.1	557	9	AY403127	Mus muscu
39	890	53.0	1603	3	CR590249	full-leng
40	890	52.8	1494	3	CR601506	full-leng
41	887	52.7	793	7	CN037447	nm_22.2_b
42	885.5	52.5	616	7	CR606695	full-leng
43	882	52.5	1485	3	CR611713	full-leng
44	882	52.5	1563	3	CR593199	full-leng
45	880.5	52.4	1791	3	BP171264	BP171264

ALIGNMENTS

RESULT 1
AK035194
LOCUS   AK035194   2730 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region
            and neck cDNA, RIKEN full-length enriched library, clone:9430098F20
            product:hypothetical Serine proteases, trypsin family containing
            protein, full insert sequence.
ACCESSION   AK035194
VERSION   AK035194.1 GI:26084456
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

```
REFERENCE    1
  AUTHORS    Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    TITLE    RIKEN integrated sequence analysis (RISA) system-384-format
             sequencing Pipeline with 384 multicapillary sequencer
  JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE    20530913
    PUBMED    11076861
REFERENCE    4
  AUTHORS    The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
    TITLE    Functional annotation of a full-length mouse cDNA collection
  JOURNAL    Nature 409, 685-690 (2001)
REFERENCE    5
  AUTHORS    The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
    TITLE    Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
REFERENCE    6
             (bases 1 to 2730)
  AUTHORS    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tonaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M and Hayashizaki,Y.
    TITLE    Direct Submission
  JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
             URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
  COMMENT    cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissue.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.jp/.
             URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..2730
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="PANTOM_DB:9430098F20"
                     /db_xref="taxon:10090"
                     /clone="9430098F20"
                     /tissue_type="embryonic body between diaphragm region and
                     neck"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     misc_feature    437..1817
                     /note="hypothetical Serine proteases, trypsin family
                     containing protein (Interpro|IPR001254, evidence:
                     InterPro)
                     putative"
     polyA_signal    2715..2720
                     /note="putative"
     polyA_site      2730
                     /note="putative"
ORIGIN
',
Alignment Scores:
Pred. No.:
```

```
Score:                    1528.00      Matches:            301
Percent Similarity:       96.32%       Conservative:        13
Best Local Similarity:    92.33%       Mismatches:          12
Query Match:              91.01%       Indels:               0
                          3            Gaps:                 0
```

```
US-10-617-443B-2 (1-334) x AK035194 (1-2730)

Qy    9 AlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValVal 28
         :::||||||||||||||||||||||||||||||||||:::|||||||||||||||||:|||
Db  837 TCTGGTCTCCACCAGCTGACCAGTCCCGGGTACAAGTTCAACTTCATCGCCGATGTGGTG 896

Qy   29 GluLysIleAlaProAlaValValHisIleGluIleProLeuPheGly 48
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  897 GAGAAGATTGCGCCAGCTGTGGTCCACATAGAGATCCCTTTGTTTGGC 956

Qy   49 ArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlnAlaGlyLeuIleIle 68
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db  957 CGGAATGTGCCCCTGTCCAGTGGCTCCGGCTTCATCATGTCACAGGCCGGTTTGATCGTC 1016

Qy   69 ThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyAlaArgGlnLeuLysVal 88
         |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db 1017 ACCAACGCCCACGTGGTCTCCAGCAACTCGGCGGCACCTGGAGCCAGGCAGCTGAAGGTG 1076

Qy   89 GlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysSerAspAsp 108
         |||||||||||||||||||||||||:::|||||||||||||||||||||
Db 1077 CAGCTGCAGAATGGGGATTCGTATGAGGCCACCATCAAGTCAGATGAC 1136

Qy  109 IleAlaThrIleLysHisProLysGlyLysLeuProValLeuLeuLeuGlyHisSer 128
         |||||||||||||||||||||||||||||||||||||||||||||||||::: 
Db 1137 ATTGCCACGATTAAACACCCCAAGGGAAAAAAGCTCCCTGTGTTGCTGCTGGGTCACTCA 1196

Qy  129 AlaAsnLeuArgProGlyGluPheValLysAlaAlaLeuAlaProGlyAlaArgGlnAsn 148
         |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db 1197 GCAAACCTGCGCCCTGGAGTTCGTGGGCCATCGGGCAGCCCCTTTGCCCTGCCAGAAC 1256

Qy  149 ThrValThrProLeuSerThrAlaGlnAsnGlyArgGluLeuGlyTyrLeuArg 168
         |||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db 1257 ACCGGTGACAACGGCGCATTGTCAGCACTCCCCAGCGTGGCAAGGAGCTGGGTCTCCGG 1316

Qy  169 AspSerAspMetAspTyrIleGlnThrAspAlaIleIleLeuAsnTyrGlyAsnSerGlyGly 188
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1317 GACTCAGACATGGACTATATCCAGACCGATGCCATCATCCTCAACTACGGGAACTCAGGAGGA 1376

Qy  189 ProLeuValAsnLeuAspGlyGlyValAlaIleGlyIleLeuAsnThrLeuLysValThrAlaGly 208
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1377 CCCCTGGTGAACCTGGATGGCGGAGTCGCATCGGCCATCAACACGCTCAAGGTTGCAGCTGGC 1436

Qy  209 IleSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGlnAsnAspLys 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||::: 
Db 1437 ATCTCCTTTGCCATCCCCTCAGATCGCATCACACGCTTCCTCTGAGTTCCAAAACAAG 1496

Qy  229 GlnIleLysAspTrpPheLysArgPheIleGlyTyrIleArgMetArgThrIleThrProSer 248
         :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1497 CATGTGAAAGACTGGAAGAACGGCTTCATTGGCATCCGGATGCGGACCATCACGCCAAGT 1556

Qy  249 LeuValAspGluGlnLeuLysAlaSerArgAsnProAspPheProGluValSerSerGlyIleTyr 268
         |||||||||||||||||||||||||||||||||||||||||||||||::: 
Db 1557 TTGGTGGAGGAACTGAAGGCCCCCAACCCAGACTTTCCAGGTCGTCACCAGCGGGAATATAT 1616

Qy  269 ValGlnGluValAlaProAsnSerProSerGlnArgGlyGlyIleGlnIleAspProValAsnIle 288
         |||||||||||||||||||||||||||||||||||||||||||||||||||||::: 
Db 1617 GTTCAAGAGGTGGTTCCCAATTCACCTTCTCAGAGAGGAGGCATCCAAGATCGGACATC 1676

Qy  289 IleValLysValAsnGlyArgProLeuValAspSerSerGlyLeuGlnGlnAlaValLeu 308
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||::: 
Db 1677 ATCGTCAAAGTCAATGGCCGCCCCCTGGTCGACGATTCCAGGCAGGCTGCAGGCCTG 1736

Qy  309 ThrGluSerProLeuLeuGluValArgAsnAspAspLeuPheSerIle 328
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1737 AACGGAGTCTTCACTCCTGCTGGAGGTGCGGCGAAATGATGATCTCCTTCTTCAGCATC 1796

Qy  329 AlaProGluValValMet 334
         ||||||||||||||||||
Db 1737 AACGGAGTCTTCACTCCTGCTGGAGGTGCGGCGAAATGATGATCTCCTTCTTCAGCATC 329
```

```
8.38e-174    Length:    2730
```

```
Db    ||||||||||||||||||||||||
RESULT 2
Db    BQ675502
      BQ675502
LOCUS      BQ675502        941 bp    mRNA    linear    EST 15-JUL-2002
DEFINITION AGENCOURT_8191926 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6256985
           5', mRNA sequence.
ACCESSION  BQ675502
VERSION    BQ675502.1  GI:21786336
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 941)
  AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
  TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL  Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM2410 row: i column: 18
           High quality sequence stop: 605.
FEATURES             Location/Qualifiers
     source          1..941
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6256985"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /clone_lib="NIH_MGC 102"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
ORIGIN

US-10-617-443B-2 (1-334) x BQ675502 (1-941)

Alignment Scores:
Pred. No.:
Score:         6.61e-143
Score:         1268.00        Length:        941
Percent Similarity: 95.60%    Matches:       255
Best Local Similarity: 93.41% Conservative:  6
Query Match:   75.52%         Mismatches:    10
DB:            5              Indels:        2
                             Gaps:          0

Qy    16 SerProArgTyrLysPheAsnPheIleAlaAspValValGluLysIleAlaProAlaVal 35
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2 AGCCCGCCGTACAAGTTCAACTTCATTGCTGACGTGGTGGAGAAGATCGCACCAGCCGTG 61

Qy    36 ValHisIleGluLeuPheLeuArgHisProLeuProLeuSerSer 55
          ||||||||||||||||||||||||||||||||||||||||||||
Db    62 GTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTGGCCCGAACCTGTCCTGTCCAGC 121

Qy    56 GlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValValSer 75
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 GGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACCAATGCCCACGTGGTGTCC 181

Qy    76 SerAsnSerAlaAlaProGlyArgGlnGlnLeuGlnLeuGlnAsnGlyAspSer 95
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   182 AGCAACAGTGCTGCCCCCGGGCAGGCAGCCAGCTTCAAGGTGCCTACAGAGAATGGGGACTCC 241

Qy    96 TyrGluAlaThrIleLysAspIleAspLysLysGlySerAspIleAlaThrIleLysHis 115
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 TATGAGGCCACCATCAAAGACATCGACAAGAAGGGTTCGGACATTGCCACCATCAAGATCCAT 301

Qy   116 ProLysGlyLysProLeuProValLeuLeuGlyHisSerAlaAspLeuArgProGlyGlu 135
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 CCCAAGAAAAAGCTCCCTGTGTTGTTGGGTCACTCGGCCGACCTTCGCCCTGGGGAG 361

Qy   136 PheValAlaIleAspIleGlySerProPheAlaLeuGluGlnAsnThrValThrThrGlyIleVal 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 TTTGTGGTGGCCATCGGACAGTCCTTCGCCCTTACAGAACACAGTGACAACGGGCATCGTC 421

Qy   156 SerThrArgAlaGlnArgGluGluGlyLeuGlyLeuArgAspSerAspMetAspTyrIle 175
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 AGCACCCGGGCCCAGCGGGAGGAGGGGCTGGGCCTCCGGGACTCCGACATGGACTACATC 481

Qy   176 GlnThrAspAlaIleLeuAsnTyrGlyAsnSerGlyValGlyProLeuValAsnLeuAspGly 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 CAGACGGATGCCATCCTCAACTACGGGGAACTCCGGGGGCGGACCACTGGTTGCAACCTGGATGGC 541

Qy   196 GluValIleGlyIleLeuAsnThrLeuLysValThrAlaGlyIleLeuSerPheAlaIleProSer 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 GAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATCCTCTCCTTTGCCATCCCCTCA 601

Qy   216 AspArgIleThrArgPhe-LeuThrArgGluPheGlnAspLysGlnIleIleLysAspTrpLysGly 235
          ||||||||||||||||||||||||||||||||||||||||||::||||::||||||||||
Db   602 GACCGCATCACACGGTTCCCTCACAGAGTTCCAAGACAAGCAGATCAAAGACTGGAAAAA 661

Qy   235 sArgPheIleGlyIleLeuArgMetArgThrIleThrProSerLeuValAspGluLeuLysAl 255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db   662 GCCCTTCATCGGCATACGGATCCGGACGATCACACCAAGCCCTTGTGGATGAACTTGAAGG 721

Qy   255 aSerAsnProAspPheProGluValSerSerGlyIleThrTyrValGlnGluValAlaProAs 275
          |||||||||||||||||||||||::|||||:::||||||||||||||||||||||||||
Db   722 CANCAACCCCGACTTCCCAGAAGGTTATCAGTGGACACATTATCGCGCAACAGGTTGCCCCGAA 781

Qy   275 nSerPro-SerGlnArgGlyGlyIleGlnAspGly 286
          ||||||  ||||:::||||||||||||::||||||
Db   782 TTCACCTTTTCCGAAAAGGGGCCCATCCCAGATGGG 816
```

```
Db    ||||||||||||||||||||||
RESULT 3
Db    CD635470/c
      CD635470
LOCUS      CD635470/c       810 bp    mRNA    linear    EST 12-JAN-2004
DEFINITION 561006633I1 FLP Homo sapiens cDNA, mRNA sequence.
ACCESSION  CD635470
VERSION    CD635470.1  GI:40283737
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 810)
  AUTHORS  Fu,G.K., Wang,J.T., Yang,J., Au-Young,J. and Stuve,L.L.
  TITLE    Circular rapid amplification of cDNA ends for high-throughput
           extension cloning of partial genes
  JOURNAL  Genomics 84 (1), 205-210 (2004)
COMMENT    Contact: Fu GK
           Incyte Genomics, Inc.
           3160 Porter Dr., Palo Alto, CA 94304, USA
           Tel: 650454102
           Email: gfu@incyte.com.
FEATURES             Location/Qualifiers
     source          1..810
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="FLP"
                     /note="Vector: pDrive Cloning Vector"
ORIGIN

Alignment Scores:
```

Pred. No. :　　　　1.51e-141
Score:　　　　　　1256.00
Percent Similarity:　99.22%
Best Local Similarity:　99.22%
Query Match:　　　74.81%

Length:　810
Matches:　255
Conservative:　0
Mismatches:　0
Indels:　2
Gaps:　0

US-10-617-443B-2 (1-334) x CD635470 (1-810)

```
Qy    79 AlaAlaProGlyAlaArgGlnGlnLeuLeuValGlnLeuGluGlyAsp--SerTyrGluI 98
            ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db   810 GCTGCCCCGGGCAGCAGCAGCTCAAGGTGCCAGTCACAGAATGGGAGACTTCCTATGAGGC 751

Qy    98 aThr-IleLysAspIleAspGlyLysLysSerAspIleAlaIleAlaThrIleLysIleHisProLysL 118
            | ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   750 CACCCATCAAAGACATCGACGGGAAGAAGTCGGACATTGCCACCATCAAGATCCATCCCAAGA 691

Qy   118 ysLysLeuProValLeuLeuLeuLeuGlyHisSerAlaAlaAspLeuHisGluPheValV 138
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   690 AAAAGCTCCCTGTTCTGTTGTTGTTGGGTCACTCGGCCGCCGACCTGCACGAGTTTGTGG 631

Qy   138 alAlaIleGlyLysSerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrA 158
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   630 TGGCCATCGGCAGTCCCTTCGCCCTTACAGAACACAGTGACAACCGGCATCGTCAGCACTG 571

Qy   158 laGlnArgGluGlyArgGluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrA 178
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570 CCCAGCGGGAGGGGCGAGGGAGCTGGGCCTCCGGGACTCCGACATGGACTACATCCAGACGG 511

Qy   178 spAlaIleIleGluAsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValI 198
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510 ATGGCCATCATCAACTACGGCAACTCCGGGGGGCCACTGGTGAACCTGGATGGAGAGGTCA 451

Qy   198 leGlyIleAsnThrLeuLysValThrAlaGlyGlyIleSerPheAlaIleProSerPheAspAlaArgI 218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   450 TTGGCATCAACACGCTCAAGGTCACGGCTGGGGGGCATCTCCTTTGCCATCCCCAGCACCCGA 391

Qy   218 leThrThrArgPheLeuThrGlnPheGlnAsnProLysGlnIleIleGluValAlaGlyArgPheI 238
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   390 TCACACGGTTCCTCACAGAGTTCCAAAGACAAAGCAGATCATCGAAGTAGCAGGGCGCTTCA 331

Qy   238 leGlyIleLeuArgMetArgThrIleThrProSerLeuValAlaSpGlyLeuValAlaSerAsnP 258
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   330 TCGGCCATACGGATGCGGACGATCACACCAAGCCTGGTGGATGGAGCTGAAGGCCAGCAAC 271

Qy   258 roAspPheProGluValSerSerGlyTyrValGlnGluValAlaAlaProAsnSerProS 278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270 CGGACTTCCCAGAGGTCAGCAGTGGAATTTATGTGCAAGAGGTTGCCGCCAATTCACCTT 211

Qy   278 erGlnArgGlyGlyGlyIleGluGlnAspIleIleGluValLysValAsnGlyAlaArgProLeuV 298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210 CTCAGAGAGGCGGCGGCCATCCAAAGATGTGACATCATCGTCAAGGTGACCGGGCGTCCTAG 151

Qy   298 alAspSerSerGluLeuGlnGluAlaValAlaLeuThrGluSerProLeuLeuLeuValA 318
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150 TGGACTCGAGTGAGCTGCAGGAGGCCGTGGCACTGACGGAGTCTCCTCTACTGGAGGTGC 91

Qy   318 rgArgGlyAsnAspAspLeuLeuPheSerIleAlaProGluValVal 333
            |||||||||||||||||||||||||||||||||||||||||||
Db    90 GGCGCGGGGGAACGACGACCTCCTTCTTCAGCATCGCCACCTGAGGTGGTC 44
```

Alignment Scores:
Pred. No. :　　　　3.42e-136
Score:　　　　　　1216.50
Percent Similarity:　85.20%
Best Local Similarity:　72.51%
Query Match:　　　72.45%

Length:　1854
Matches:　240
Conservative:　42
Mismatches:　46
Indels:　4
Gaps:　1

US-10-617-443B-2 (1-334) x CR731516 (1-1854)

```
Qy     7 AlaSerAlaGlyValLeuHisGlnLeuSerSerProAlaGlyTyrLysProPheAsnPheIleAlaAsp 26
            : :      ::|     ::|        |||:      ::   |||      ||||||
Db   571 AGCGGCGTCAGGTTCCTCGGCCGCCGCCTTCAGCCCTTCGCTCTACAGAGTTTAACTTCATCGCCGAC 630

Qy    27 ValValGluLysIleAlaProAlaValAlaValHisIleGluLeuPheLeuArgHisProLeu 46
            :||     ::  :||   |||      |||:||||||||    ::|||   |||||
Db   631 GTGGTGGAGAAGATAGCGCCCGCCGTGGTCCACATCGAGCTCTTCCTGAGGCACCCCCTC 690

Qy    47 PheGlyArgAsnValProLeuSerSerGlyGlyPheIleMetSerGluAlaGlyLeu 66
            ||||||     |||      :|||      |||::|         |||   |||
Db   691 TTTGGTCGGACTGTCCCCCTGTCAAGTGGATCCGGCTTCCTGGTGTCGGAGAACGGGCTG 750

Qy    67 IleIleThrAsnAlaHisValValSerSerAsnSerAlaAlaAlaProGlyArgGlnLeu 86
            ||      :|        |||      |||      ||||||      |||   |||
Db   751 ATCGTCACCAACGCCCACGTGGTCTCCAGCAACTCAGCAGCCGCTCAGCAGCACCTC 810

Qy    87 LysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysLysAspIleAspGlyLysLys 106
            |||   ||   |:|      :|   :||   |||   |||      ||||||   :|||
Db   811 AAGGTCCAGATACGGAACGGGGACAGCTACGAAGCCACCATCAAGAAAGACATTGACGGCAAAAA 870

Qy   107 SerAspIleAlaThrIleLysIleHisProLysLysLysLeuLeuLeuGly 126
            ||||||   |||   |||:||||||         ::|||   |||
Db   871 TCCGACATCGCCACCATCAAGGTCCACCCGAGGAAAGTCCCGGTCCTGTTTCTGGGC 930

Qy   127 HisSerAlaAlaAspLeuArgProGlyGluPheValValAlaAlaIleGlySerProPheAlaLeu 146
            |||      ||||||         |||   ||||||      |||||||||||:||   :||
Db   931 CAGAGCGCCGCGACCTGAGACCCGGGGAGTTTGTGGTCGCAGCCATCGGCTCCCC 990

Qy   147 GlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArgGluLeuGly 166
            |||:||   |||   |||||||||   :||||||||||   ||:|||      |||
Db   991 CAGAACACCGTCACCACCGGCATCGTCAGCACCGCTCAGCGCGAGGGGCGGGAGCTGGGC 1050

Qy   167 LeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleGluAsnTyrGlyAsnSer 186
            ||||||      |||   |||:||:||      |||   :|||||||||:||   :||   ||:
Db  1051 CTGAGGGACACTCGGACAAGCCTACGGCCTCCAGACGGACGCGATCATCAACTACGGCAACTCT 1110

Qy   187 GlyGlyProLeuValAsnLeuAspGlyGluValIleGlyIleIleAsnThrLeuLysValThr 206
            |||||||||   ||||||||||||      |||:||:||||||   |||:|||||||||   :|
Db  1111 GGAGGGCCCCTGGTCAACTTGGACGGAGAGGTGATCGGGATCAACACCCTGAAGTGCG 1170

Qy   207 AlaGlyIleSerProPheAlaIleProSerPheAspAlaArgIleThrArgProPheLeuThr----- 223
            |||||||      :||||||::|      |||   |||||||   ||:||:||||||
Db  1171 GCGGGCATCTCCCCTGCTATCCCGTCCGACGCAGGATTGCCCGGTTCCTGACCCACCCC 1230

Qy   224 GluPheGlnAsnAspLysGlnIleIleGluValLysValAsnGlyAlaArgPheIleGlyIleArgMetArg 243
            |||||||||||||   :||||||||||||   |||:||   |||      |||||:   |||
Db  1231 GAGTTCCAGAAC... 243
```

RESULT 5

LOCUS   BQ080316   890 bp   mRNA   linear   EST 16-AUG-2002
DEFINITION AGENCOURT_8058818 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6213159
            5', mRNA sequence.
ACCESSION   BQ080316
VERSION   BQ080316.1 GI:22272324
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 890)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2378 row: g column: 16
            High quality sequence stop: 578.
FEATURES   Location/Qualifiers
     source   1..890
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6213159"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Qy   1231 CAGAGGCCACGGCAAAGATGCCAAGTCCGCAAAGAAGAAGCGGCTTCATCGGGATCAGGATCGCTG 1290
           ::: ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1291 ACCGTCACGCCAGGGTTGATGGAGGAGGAGGAGGATCCAGACTTCCTGACATC 1291

Qy   244 ThrIleThrProSerLeuValAspGluGluLeuHisArgAspGlnAspArgLeu 263
           ||:::||||||||||||||
Db   1351 SerSerGlyIleTyrValGlnGluValAlaAlaProAsnSerProSerGlnArgGlyGlyIle 263

Qy   264 SerSerGlyIleTyrValGlnGluValAlaAlaProAsnSerProSerGlnArgGlyGlyIle 283
           :||||||||||||||||||||||||||||||:::
Db   1351 AGCGGCGGGATTTACGTACGCACCGAGTGTGCCCCTGTCCGCCCGCAGATAAAGGGGCATC 1350

Qy   284 GlnAspGlyAspIleIleValLeuLysValSerGlyAsnGlyArgProLeuValAspSerProSerGlnLeu 303
           ||||||||||||||||||||
Db   1411 AAAGGAGGGGCAGAGTCCTGGTCAAGCTCAAGCTCCACACCCCGACCTG 303

Qy   304 GlnGluLeuValLeuLeuThrGluSerProLeuLeuLeuGluValValArgArgGlyLysAsnAspAsp 323
           ||||::: |||||||||
Db   1470 CAGGGGGGCGGCTGCAGGAGGAGGCGTGCTGCTGGAGGTGCGCAGGGGCAACGACGAC 1529

Qy   324 LeuLeuPheSerIleAlaProGluValValMet 334
           ||||||||||||||||||||
Db   1530 CTGCTCTTCAACATCGCCGGAGACCTCATCCTG 1562

RESULT 6

LOCUS   CD635467   724 bp   mRNA   linear   EST 12-JAN-2004
DEFINITION 56056242H1 FLP Homo sapiens cDNA, mRNA sequence.
ACCESSION   CD635467
VERSION   CD635467.1 GI:40283734
KEYWORDS   EST.

US-10-617-443B-2 (1-334) x BQ080316 (1-890)

Best Local Similarity: 91.10%
Query Match: 70.94%
DB: 5

Mismatches: 9
Indels: 12
Gaps: 2

Qy   16 SerProArgTyrLysPheAsnPheIleAlaLeuProValValAlaGluGluLysIleAlaProAlaVal 35
          ||||||||||||||||||||||||||||||||||||||
Db   2 AGCCCGCGCTACAAGTTCAACTTCATTGCTCTGACGTGGTGAGGAGAGATGCAACCAGCGTG 61

Qy   36 ValHisIleGluLeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSer 55
          ||||||||||||||||||||||||||||||||||||
Db   62 GTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGGAACGTGCCCCTGTCCAGC 121

Qy   56 GlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValValSer 75
          |||||||||||||||||||||||||||||||
Db   122 GGTTCTGGCTTCATCATGTCAGAGGCGGGGCTGATCATCACCAATGCCCACGTGGTGTCC 181

Qy   76 SerAsnSerAlaAlaProLysValArgGlnLeuGlnLeuLysValGlnLeuAsnGlyAspSer 95
          |||||||||||||||||||||||||||||||||||
Db   182 AGCAACAGTGCTGCTCCCAAGGTGCAGCTACAGAATGGGGACTCC 241

Qy   96 TyrGluAlaThrIleLysAspIleAspLysLysSerAspIleAlaThrIleLysIleHis 115
          |||||||||||||||||||||||||||||||||||||
Db   242 TATGAGGCCACCATCAAAGACATCGACAAGAAGAGTCGGACATTGCCACCATCAAGATCCAT 301

Qy   116 ProLysGlyLysGlnProValLeuLeuLeuGlyHisSerAlaSerGluLeuArgProGlyGlu 135
          ||||||||||||||||||||||||||||||||||
Db   302 CCCAAGAAAGAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCCAGCGAGCTGCGGGGAG 361

Qy   136 PheValValAlaIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleVal 155
          ||||||||||||||||||||||||||||||||||
Db   362 TTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACAGTGACAACGGGCATCGTC 421

Qy   156 SerThrAlaGlnArgGluGlyLeuArgGluLeuGlyLeuGluArgAspSerAspMetTyrIle 175
          ||||||||||||||||||||||||||||||||
Db   422 AGCACTGCCCAGCGGGAGGGGCTGCGGGAGCTGGGCCTCGGGAGGACTCGGACTACATC 481

Qy   176 GlnThrAspAlaAlaIleIleGlySerGlyAsnSerGlyGlyProLeuValAsnLeuAspGly 195
          ||||||||||||||||||||||||||||||||
Db   482 CAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGGCCACTGGTGAACCTGGATGGC 541

Qy   196 GluValIleGlyIleLeuAsnThrLeuValValValThrAlaGlyIleSerPheAlaIleProSer 215
          |||||||||||||||||||||||||||||||||||
Db   542 GAGGTCATTGGCATCAACACGCTCAAGGTGCATCCTTTGCCAATCCCCTCA 601

Qy   216 Asp---ArgIleThrArgPheLeuThrGlnPheGlnAsnIleLeuLysAlaSerTrp-Ly 234
          ::: |||||||||||||||||||||||||||||:::|||
Db   602 GAACCCGCATCACACGGGTTCCTCACAGAGTTCAAGAGACAAGCAGATCAAAGAACTGGGAA 661

Qy   234 SerLysArg-PheIleGlyLeuArgMetArgThrIleThr-ProSer-LeuValAspGluGlu 253
          ||:||||||| |||||||||||||||||||| ||||||
Db   662 GGAAGCGGCTTCATCGGGCCGATACGGACGATCACCACCAAAGCCCTGGTGGGATGAGCT 721

Qy   253 uLysAlaSer-AsnProAspPheProGluVal--SerSerGlyIleTyrVal-GlnGluV 272
          |||||||||||||| |||||||||||| ::: ||||||||||||||||||
Db   722 GAAGGCCAGCAAAACCCCGACTTCCCAGAGGTTTAACAGTGGGAAATTTATGTGCAAGAGG 781

Qy   272 alAla-ProAsnSerProSerGlnArgGly-GlyIleGlnAspGlyAsp---IleIleVa 290
          ||||||||||||||||||||||||||| |:::|||||||||| ||||||
Db   782 TTGCCCCCCAATTCACTTTCTCAAAGAGGCCGGCATCCAAAATGGTGAACATCATCCGTC 841

Qy   290 lLysValAsnGlyArg 295
          ::|||||||||||||
Db   842 AAGGGTCAACGGCGGG 857

SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 724)
AUTHORS   Fu,G.K., Wang,J.T., Yang,J., Au-Young,J. and Stuve,L.L.
TITLE   Circular rapid amplification of cDNA ends for high-throughput
    extension cloning of partial genes
JOURNAL   Genomics 84 (1), 205-210 (2004)
COMMENT   Incyte Genomics, Inc.
    Contact: Fu GK
    3160 Porter Dr., Palo Alto, CA 94304, USA
    Tel: 6508454102
    Email: gfu@incyte.com.
FEATURES   Location/Qualifiers
source   1..724
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone_lib="FLP"
    /note="Vector: pDrive Cloning Vector"

ORIGIN

US-10-617-443B-2 (1-334) x CD635467 (1-724)

Alignment Scores:
Pred. No.:   6.76e-132   Length: 724
Score: 1176.00   Matches: 237
Percent Similarity: 98.35%   Conservative: 1
Best Local Similarity: 97.93%   Mismatches: 3
Query Match: 70.04%   Indels: 1
                                         Gaps: 0

```
Qy   80  AlaProGlyArgGlnGlnLeuLys-ValGlnLeuGlnAsnGlyAspSerTyrGluAlaTh  99
Db    1  GCTCGAGGCAGGCAGCAGCTCAAGGNTGCAGCTACAGAAATGGGGAGACTCCTATGAGGGCCAC  60

Qy   99  rIleLysAAspIleAsnGlyLysSerAspIleAlaThrIleLysIleLeuHisProLysLysGly  119
Db   61  CATCAAAGACATCGACAAGAAAGTCGGACAATTGCCACCATCAAGATCAATCCCAAGGAAAA  120

Qy  119  sLeuProValLeuLeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValValAl  139
Db  121  GCTCCCTGTTGTTGTTGCTGGGTCACTCGGCCGACCTGCGGCCTTGGGGAGTTTGTGGTGGC  180

Qy  139  aIleGlySerProPheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGl  159
Db  181  CATCGGCAGTCCCTTCGCCCTACAGAACACAGTGACAACGGGCATCGTCAGCACTGCCCCA  240

Qy  159  nArgGlyLysGlyArgGluLysGlyLeuArgGlnAspSerMetAspTyrArgIleGlnThrAspAl  179
Db  241  GCCGGGAGGGGCAGGGAGCTGGGGCCTCCGGGACTCCGACATGGACTACAGACGGATGC  300

Qy  179  aIleIleAsnThrTyrGlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleGl  199
Db  301  CATCATCAACTACGGGAACTCCGGGGACCACTGTGGAACCTGGATGGCGGAGGTCATTGG  360

Qy  199  yIleAsnThrLeuLysValThrAlaGlyIleLeuSerPheAlaIleProSerArgAspArgIleTh  219
Db  361  CATCAACACGCTCAAAGTCACGGCCGGCATCCTTTGCCATCCCCTTCGACCGCGGGCATCAC  420

Qy  219  rArgPheLeuThrGlnPheGlnIleAspLysAspThrProLysLysArgPheIleGl  239
Db  421  ACGCGTTCCTCACAGAGTTCCAAGACAACAAGACTGAAGAAGCGCGTTCATCGGG  480

Qy  239  yIleArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAs  259
Db  481  CATACGGATGCGGACGATCACCACCAAGCCTGGTGGATGAGCTGAAGGCCAGCAACCCCGA  540

Qy  259  pPheProGluValSerSerGlyIleuTyrvalGlnGluValAlaProAsnSerProSerGl  279
Db  541  CTTCCCAGAGGTGTCAGCAGTGGAATTTATGTGCAAGAGGGTTGCGCCGAATTCACCTTCTCA  600
```

RESULT 7
CN787405

LOCUS   CN787405   757 bp   mRNA   linear   EST 26-MAY-2004
DEFINITION   4121511 BARC_8BOV Bos taurus cDNA clone 8BOV_34P04 5', mRNA
    sequence.
ACCESSION   CN787405
VERSION   CN787405.1   GI:47683385
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
ORGANISM   Bos taurus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
    Bovinae; Bos.
REFERENCE   1 (bases 1 to 757)
AUTHORS   Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassell,C.P. and
    Matukumalli,L.K.
TITLE   Construction and Analysis of a cDNA Library Generated From
    Intestinal Muscle and Epithelial Tissues of Holstein Cattle
JOURNAL   Unpublished (2004)
COMMENT   Contact: Richard G. Baumann
    Bovine Functional Genomics Lab
    ANRI
    BLDG 162: BARC-EAST, Beltsville, MD 20705, USA
    Tel: 3015048604
    Fax: 3015048744
    Email: rbaumann@anri.barc.usda.gov
    Single pass sequencing. Bases called and trimmed with phred
    0.000925 using options -trim alt '' -trim fastaVector identified by
    cross_match using options -minmatch 12 -minscore 12
    Plate: 34 row: P column: 04
    Seq primer: CCTATTTAGGTGACACTATAGAAC
    High quality sequence stop: 757.
FEATURES   Location/Qualifiers
source   1..757
    /organism="Bos taurus"
    /mol_type="mRNA"
    /strain="Holstein"
    /db_xref="taxon:9913"
    /clone="8BOV_34P04"
    /sex="Female"
    /tissue_type="Epithelial, Muscle"
    /dev_stage="Lactating, Neonatal"
    /lab_host="DH10B TonA"
    /clone_lib="BARC 8BOV"
    /note="Organ: Intestine; Vector: pCMVsport6.1; Site_1:
    NotI; Site_2: EcoRI; Normalized cow cDNA intestinal
    library in pCMVsport6.1, constructed from equimolar mRNA
    pools derived from 5 sources, 4 lactating intestinal, 1
    neonatal intestinal 4/5 Lactating, Proximal Duodenum,
    Jejunum, Distal Ileum, Colon, 1/5 Neonatal, Proximal
    Duodenum, Jejunum, Jejunum, Distal Ileum"

ORIGIN

Alignment Scores:
Pred. No.:   1.48e-129   Length: 757
Score: 1157.00   Matches: 233
Percent Similarity: 94.84%   Conservative: 6
Best Local Similarity: 92.46%   Mismatches: 13
Query Match: 68.91%   Indels: 1
                                         Gaps: 0

```
Qy  279  nArgGlyGlyIleGlnAspGlyAspSerIleLeuValLysValValAsnGlyAlaProLeuValAs  299
Db  601  GAGAGGGCGGCCATCCAAGATGGGTGACATCATCGTCAAGGTCGTCAACGGGGCTCCTCTAGTGA  660

Qy  299  pSerSerGluLeuGlnGluAlaValLeuThrGluSerProLeuLeuLeuLeuValArgAr  319
Db  661  CTCGGAGTGAGCTGCAGGAGGCGGTGCTGACCGAGTCTCCTCTTACTGAGGTGGGCGGG  720

Qy  319  gGly  320
Db  721  GGGA  724
```

US-10-617-443B-2 (1-334) x CN787405 (1-757)

```
Qy    66 LeuIleIleThrArgAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGln 85
         ||||||||||||||||||||||||       :::          ||||||||||||||
Db     2 CTGATTGTCACCAACGCACCATGTGGTGTCCAGCACCACCAACGTCTCGGGCCGGCAGCAG 61

Qy    86 LeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleLysGlyAspIleAspLys 105
         ||||||||||||||||||||::::          |||||||||||||||||||||
Db    62 CTCAAAGTGCAGCTGCAGAATGGGTGACACTTACGAGGCCACCATCAAGGGACATCGACAAG 121

Qy   106 LysSerAspIleAlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuLeu 125
         |||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db   122 AAGTCGGACATCGCCACCATCAAGATCCACCCAAGAAAAAAGCTCCCTGCCTGCTG 181

Qy   126 GlyHisSerAlaAlaAspLeuArgProGlyGluPheValAlaIleGlySerProPheAla 145
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 GGCCACTCGGCAGACCTGCGCCCGGGAGTTTGTGGCCATCGGCAGCCCCTTCGCC 241

Qy   146 LeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnAsnThrLeuLysVal 165
         ||||||||||||||||||||||||||||||||::::|||||||||||||
Db   242 CTGCAGAACACGGTGACCACCGGCATCGTGAGCACAGCCCAACGGGAGCTG 301

Qy   166 GlyLeuArgAspSerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsn 185
         |||||||||||||||||||||||||||||||||||||||||||:::|||
Db   302 GGCCTTCGGGACTCTGACATGGACTACATTCAGACAGACGCCATCATCAATTACGGGAAC 361

Qy   186 SerGlyGlyProLeuValAsnLeuAspGlyValIleIleGlyIleAsnThrLeuLysVal 205
         |||||||||||||||||||||||||||||||||::::||||||
Db   362 TCCGGCGGCCCGCTGGTGAACCTGGACGGGGTCATTATTGGCATCAACACTCAAGGTC 421

Qy   206 ThrAlaGlyIleLeuSerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPhe 225
         ||||||||||||||||||||||||||||||||||||:::|||
Db   422 GCGGGCCGGCATCTCCTTTGCCATCCCCTCAGACCGGATCACGCGCTTCCTCAGAGTTC 481

Qy   226 GlnAspLysGlnIleIleLysAspSerPheIleGlnArgMetArgThrIle 245
         ||||||||||||||||||||||||||||||||||||||:::
Db   482 CAGGACAAGACTGCAAGACTCGAAGAGAAGGCGCTTCATCGGCATCCGGATGAGGACCATC 541

Qy   246 ThrProSerLeuValAspGluGlyIleLeuLysAlaSerAsnProAspGluProGluValSer 265
         ||||||||||||||||||||||||||||||||||||:::
Db   542 ACGGCCAAGTTTGGGTGGAGGAGCTAAAGGCCAGTAACCCAGAGGTCAGCAGT 601

Qy   266 GlyIleTyrValGlnGluValAlaLeuArgAsnAspSerThrArgGlyGlyIleIleAsp 285
         ||||||||||||||||||||||||||||||||::::
Db   602 GGGATTTATGTGCAAGAAGTTGTTCGGAATTCACCTTCTCAGAGCAGTCAGCAGT 661

Qy   286 GlyAspIleIleGluValLysValAspGlyProLeuValAspSerSerGluLeuGlnGlu 305
         ||||||||||||||||||||||||||||||||||||
Db   662 GGAGACATCATCGTCAAGGTCAACGGGCCGCGCCGGGCCCTCAGCCGAGCTGCAGGAG 721

Qy   306 AlaValLeuThrGluSerProLeuLeuLeuGluVal 317
         |||||||||||||||||||||||||||||
Db   722 GCCGTGCTGACCGAGTC-CCCCTGCTGCTGGAGGTG 756
```

Alignment Scores:
Pred. No.:   4.76e-124
Score: 1117.00   Length: 1990
Percent Similarity: 100.00%   Matches: 222
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 66.53%   Mismatches: 0
                                      Indels: 0
DB: 3   Gaps: 0

US-10-617-443B-2 (1-334) x CR604057 (1-1990)

```
Qy    10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db    10 GGTCTCCACCAGCTGAGCAGCCCCGCTACAAGTTCAACTTCATTGCCGATGTGGTGGAG 636

Qy    30 LysIleAlaProAlaValAlaHisIleGluLeuPheLeuArgHisProLeuPheGlyArg 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   577 AAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACATCCCGCTGTTTGGCCGC 696

Qy    50 AsnValProLeuSerGlySerGlyPheIleMetSerGluAlaGlyLeuIleIleThr 69
         ||||||||||||||||||||||||||||||||||||||||||
Db   637 AACGTGCCCCTGTCCAGCGGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACC 756

Qy    70 AsnAlaHisValValSerSerAsnSerAlaAlaProGlyArgGlnGlnLeuLysValGln 89
         |||||||||||||||||||||||||||||||||||||||||||||
Db   697 AATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGACGCCAGCAGCTCAAGGTGCAG 816

Qy    90 LeuGlnAsnGlyAspSerTyrGluAlaThrIleLysGlyAspIleAspLysLysSerAspIle 109
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   817 CTACAGAATGGGGACTCCTATGAGGCCACCATCAAGGGACATCGACAAGAAGTCGGACATT 876

Qy   110 AlaThrIleLysIleHisProLysLysLysLeuProValLeuLeuLeuGlyHisSerAla 129
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   877 GCCACCATCAAGATCCATCCCAAGAAAAAAGCTCCCTGTGTTGTTGCTGGGGTCACTCGGCC 936

Qy   130 AspLeuArgProGlyGluPheValAlaIleGlySerProPheAlaLeuGlnAsnThr 149
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   937 GACCTGCGGCCTGGGGAGTTTGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACA 996

Qy   150 ValThrThrGlyIleValSerThrAlaGlnAsnThrLeuLysValGlyLeuArgAspTyr 169
         ||||||||||||||||||||||||||||||||||||||||||||
Db   997 GTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGCCTGGGCCTCCGGGAC 1056

Qy   170 SerAspMetAspTyrIleGlnThrAspAlaIleIleAsnTyrGlyAsnSerGlyGlyPro 189
         ||||||||||||||||||||||||||||||||||||||||||||
Db  1057 TCCGACATGGACTACATTCAGACGGATGCCATCATCAACTACGGGAACTCCGGCGGGCCA 1116

Qy   190 LeuValAsnLeuAspGlyValIleIleGlyIleAsnThrLeuLysValThrAlaGlyIle 209
         |||||||||||||||||||||||||||||||||||||||||
Db  1117 CTGGTGAACCTGGATGGCGGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCGGCATC 1176

Qy   210 SerPheAlaIleProSerAspArgIleThrArgPheLeuThrGluPheGlnAspLysGln 229
         |||||||||||||||||||||||||||||||
```

US-10-617-443B-2 (1-334) x CO404217 (1-796)

```
Qy   1177 TCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAG 1236
          ||||||||||||
Db    230 IleLys 231

Db   1237 AtCAAA 1242
```

ORIGIN

Alignment Scores:
Pred. No.:

Score:	2.99e-119	Length:	796
Percent Similarity:	1072.50	Matches:	216
Best Local Similarity:	95.40%	Conservative:	12
Query Match:	90.38%	Mismatches:	10
	63.88%	Indels:	2
	7	Gaps:	1

DB:

RESULT 9

LOCUS   CO404217   796 bp   mRNA   linear   EST 01-JUL-2004
DEFINITION   AGENCOURT 27851724 NIH_MGC_252 Rattus norvegicus cDNA clone
             IMAGE:7307090 5', mRNA sequence.
ACCESSION   CO404217
VERSION   CO404217.1 GI:49586133
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 796)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
             Office of Cancer Genomics
             National Cancer Institute / NIH
             Bldg. 31 Rm10A07 Bethesda, MD 20892
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
             College of Wisconsin
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM15347 row: o column: 24
             High quality sequence stop: 701.
FEATURES   Location/Qualifiers
     source   1..796
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7307090"
                     /tissue_type="Ovary, strain - Norway Line3. Age8 weeks."
                     /clone_lib="NIH_MGC_252"
                     /note="Organ: ovary; Vector: pDONR 201; Site_1: NotI;
                     Site_2: NotI; RNA obtained from female overies animals at
                     8 wk_old. Tissues were snap-frozen and kept at -80C for
                     two days before RNA extraction and purification
                     (Tri-reagent method). cDNA was primed using oligo-dT
                     primer: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCC(T)25-3' and
                     cloned into the EcoRV/NotI sites of pExpress-1.
                     Size-selection >1.25kb resulted in an average insert size
                     of 1.7kb. This primary library is not normalized
                     (normalized library is NIH_MGC 252) and was constructed by
                     Express Genomics (Frederick, MD). Note: this is a NIH_MGC
                     library"
```

```
Qy     22 AsnPheIleAlaAspValAlaGlyLysGlyIleAlaProAlaValValHisIleGluLeuPhe 41
          |||||||||||||||:|||||||| ||||||  || | |||:||||| |||| |||||||||
Db      3 CACTTCATCGCCGATGTGGTGGAACAGACATCGGCCGCAGCCGTGGTCCATATAGAGCTCTTT 62

Qy     42 LeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMet 61
          ||||||||||||||| |||| |:|||| |||||||||||||||||||||||||||:::
Db     63 CTGAGGCACCCTCTGTTTGGCCGAAAATGTGCCACTGTCCAGTGGCTCGGGGTTCATCATG 122

Qy     62 SerGluAlaGlyTyrLeuIleIleThrArgAsnAlaHisAlaValSerSerAenSerArgLys 81
          ||||||||||||  || ||:|||||||| |||||| |||||||| |:::
Db    123 TCAGAAGCCGGGTTGATCGTCACCAGGAACGCCCACGCCGTGTCCA  182

Qy     82 AspIleAspProLysGlySerAspIleAlaThrArgLeuLysIleHisAlaProLysLysSerLeuPro 121
          ||||||||||||| ||| |||||||||||||||||||||||| |||| ||||||| |||||
Db    183 GACATCGACCCAAAGAGTCAGACATCGCCACCAGGCTCAAGATCCACGCCCCAAAAAAGCTCCC 242

Qy    102 SerProPheAlaLeuGlnAsnThrValThrThrGlyIleLeuValSerThrAlaGlnArgGlu 161
          ||||| ||||| ||||||||||||||||||||| |||||||||||| |||:::
Db    243 AGTCCCTTTGCCCTGCAAAACACCGTGACAACGGGCATCCTGGTGTCCACTGCCCAGCGGGAT 302

Qy    122 ValLeuLeuLeuGlyHisSerArgAlaAspLeuArgProGlyGlyPheValValAlaIleGly 141
          |||||:||||| ||| ||| || |||||| |||||||||||||||||||||||||||
Db    303 GTGTTACTGCTGGGTCACTCGCGAGCCGACCTGCGGCCTGGGGGGTTTGTTGTGGCCATCGGC 362

Qy    142 SerProPheAlaLeuGlnAsnThrValThrThrGlyIleLeuValSerThrAlaGlnArgGlu 161
          ||||| ||||| ||||||||||||||||||||| ||||||||||||  |||:::
Db    363 AGTCCCTTTGCCCTGCAAAACACCGTGACAACGGGCATCCTGGTGTCCACTGCCCAGCGGGAT 422

Qy    162 GlyArgGlyLeuGlyLeuArgAspSerAspMetAspTyrThrIleGlnThrAspAlaIleIle 181
          |||| |||:|| ||||| |||| ||| |||||||| |||||||| |||||| ||||| ||
Db    423 GGCAAAGAGCCTGGGCCTCGGGGATTCGGACATGGACTATATCCAGACAGACGCCATCATC 482

Qy    182 AsnTyrGlyAsnSerGlyGlyProLeuValAsnLeuAsnGlyGluValIleIleGluAsn 201
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    483 AATTATGGAAACTCAGGAGGCCCCCTGGTGAACCTGAATGGTGAGGTCATCGGCATCAAC 542

Qy    202 ThrLeuLysValThrAlaGlyIleLeuSerProSerArgIleThrArgPhe 221
          |||||||||||| ||||||| |||||||| |||||| ||||||||||
Db    543 ACGGCTCAAGGTGGGGGCGTGGGGTATCTCCTTTGCCATCCGGATCACGGCGTTC 602

Qy    222 LeuThrGluThrGluGlnAspLeuLysSerGlyAspTrpLysLeuValSerArgPheIleLeuArg 241
          ||||||| |||| |::|||||||||:||||| |:||||||||||||||||||:||:
Db    603 CTCTCTGAGTTCCAAGACAAGCATGTGAAAAGACTGGAAGAGGCCTTCATTGGCCATACTCCGG 662

Qy    242 MetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAspPhe 260
          || ||| |||||||||||||| ||||||||||||||||||||| |||:
Db    663 ATGCGGACAATC-ACGCCAAGTTTCGGTGAGGACTGAGACTGC---AACCCAGATTTC 715
```

RESULT 10

LOCUS   CV106197   798 bp   mRNA   linear   EST 30-AUG-2004
DEFINITION   AGENCOURT 31536328 NIH_MGC_269 Rattus norvegicus cDNA clone
             IMAGE:7460898 5', mRNA sequence.
ACCESSION   CV106197
VERSION   CV106197.1 GI:51614770
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 798)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
             Office of Cancer Genomics
             National Cancer Institute / NIH
             Bldg. 31 Rm10A07 Bethesda, MD 20892
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Josef Lazar, Dr. Howard J. Jacob, Medical
             College of Wisconsin
             cDNA Library Preparation: Open Biosystems

```
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM15731 row: 1 column: 16
          High quality sequence start: 8
          High quality sequence stop: 750.
FEATURES             Location/Qualifiers
     source          1..798
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7460898"
                     /tissue_type="whole placenta, 2 pooled"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_269"
                     /note="Organ: placenta; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; Tissue was collected from two pooled
                     placentas from the 21st day of pregnancy, 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was cloned into the Not I and EcoRV
                     sites of pExpress-1. Library was size-selected for >1.25
                     kb fragments for an average insert size of 2.05 kb. A
                     normalized version of this library is also available
                     (NIH_MGC 270). Library was constructed by Open Biosystems
                     (Huntsville, AL). Note: this is a Mammalian Gene
                     Collection library"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                    7.5e-118
Score:              1061.00          Length:         798
Percent Similarity: 91.12%           Matches:        219
Best Local Similarity: 84.56%        Conservative:    17
Query Match:        63.19%           Mismatches:      20
DB:                      7          Indels:           6
                                    Gaps:             2
```

```
US-10-617-443B-2 (1-334) x CV106197 (1-798)

Qy     40 LeuPheLeuArgHisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPhe 59
          ::|||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db     32 ATCTTTCTGAGGCACCCTCTGTTTGGCGAAATGTGCCACTGTCCAGTGGCGTCGGGCTTC 91

Qy     60 IleMetSerGluAlaGlyLeuIleIleThrAsnAlaHisValValSerSerAsnSerAla 79
          ||:||||||||||||||||||||||||||||||||||||||||||||||||:||:::
Db     92 ATCATGTCAGAAGCCGGCCTTGATCGTCACCAACGCCCACGTGGTCTCCAGCTCCAACACC 151

Qy     80 AlaProGlyArgGlnGlnLeuLysValGlnLeuAsnGlyAspSerTyrGluAlaThr 99
          :||||||||||||||||||:||:||||:|||||||||||||||||::||||||||||
Db    152 GCCTCAGGTCGGCAGCCTGCAAGGTGCAGAATGGGGATGGCTTATTGAGGCCACC 211

Qy    100 IleLysAspIleAspGlyLysSerAspIleAlaThrIleLysIleHisProLysLysLys 119
          :||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db    212 ATCAAGGACATCGACGGAAAGTCAGACATCGCCACCATCTTAATCCACCCCAACAAAAAG 271

Qy    120 LeuProValLeuLeuLeuGlyHisSerAlaAspLeuArgProGlyPheValValAla 139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 CTCCCCGTGTTACTCCTGGGTCACTCGGCAGACCTGCGGCCTGGAGTTTGTGGTGGCC 331

Qy    140 IleGlySerProPheAlaLeuGlnAsnThrValThrGlyIleValSerThrAlaGln 159
          ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db    332 ATCGGCAGTCCCTTTGCCCTGCAAAACACCGTGACAACGGGCATCGTCAGCACTGCCCAG 391

Qy    160 ArgGlyArgGluLeuGlyLeuArgAspSerAspTyrIleGlnThrAspAla 179
          ||:||||||||:||:|||||||||||||||||||||||||||||||||||||||
Db    392 CGGGGATGGCAAAAGAGCCTGGGGACCATGGACTATATCCAGACAGACGCC 451

Qy    180 IleIleAsnTyrGlyValAsnSerGlyGlyProLeuValAsnProGlyValIleGly 199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    452 ATCATCAATTATGGAAACTCAGGAGGCCCCTTGGTTGAACCTGGAGGTCATCGGC 511

Qy    200 IleAsnThrLeuLysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThr 219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          219
```

```
RESULT 11

LOCUS       AY406047                973 bp    DNA     linear    GSS 15-DEC-2003
DEFINITION  Homo sapiens PRSS11 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY406047
VERSION     AY406047.1  GI:39762021
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 973)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..973
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1..>973
                     /gene="PRSS11"
                     /locus_tag="HCM2427"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                    1.69e-114
Score:              1034.50          Length:         973
Percent Similarity: 80.43%           Matches:        203
Best Local Similarity: 62.08%        Conservative:    60
Query Match:        61.61%           Mismatches:      55
DB:                      9          Indels:           5
                                    Gaps:             2
```

```
US-10-617-443B-2 (1-334) x AY406047 (1-973)

Qy    512 ATCAACACAGCTCAAGGTGGGCGGCCGCTGGTATCTCCTTTGCCATCCCCTCGGATCGCATCACG 571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    512 CGCCTTCCTCTCTCTGAGTTCCAAGACAACGACATGTGAAGACTGGAAGAAGAGCGGCTTCATTGGC 571

Qy    220 ArgPheLeuThrGluPheGlnAspLysGlnIleLysAspThrLysSerArgPheLeuGly 239
          |||||||:|||||||||||||:|||||||||||||||:::|||||||||||||||||||
Db    572 CGCCTTCCTCTCTCTGAGTTCCAAGACAACGACATGTGAAGACTGGAAGAAGAGCGGCTTCATTGGC 631

Qy    240 IleArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAsp 259
          :||||||||||:|:||||||||||||||||||||||||||||||||::|||||||||||
Db    632 ATCCGGATGCGAACCATCACCCCAAGTTTGGTGGACGAACTGAAGACTGCCAACCCAGAT 691

Qy    260 PheProGluValSerGlyIleTyrValGlnGluValAlaProAsnSerProSerGln 279
          |||||||||||||:||:||||||||||||||||||||:::||||||||||||||||||
Db    692 TTTCCA-GCAGTGAGCAGTGGATA-TATGTTCA-GAGGTCGGTTCCCCACTCACTTCTCAGAG 748

Qy    280 ArgGlyGlyIleGluGlnAspGlyAspIleIleIleValLysValAsnGlyArgProLeuVal 298
          ||||||||||:::|||||||||||||||||:::
Db    749 AGA---GGCATCCAGATGCGA------CTCAATCAAGTCATCGGACTCCCTGTAGTT 796
```

```
US-10-617-443B-2 (1-334) x AY406047 (1-973)

Qy     10 GlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPheIleAlaAspValValGlu 29
          |||      :::||||||||||||||||||||:||:|||||||||||||||||||||||||
Db      2 GGGCAGGAGGAATCCCAACAGTTTTGCGCCATAAAATATAACTTTATCGCGGACGTGGTGGAG 61
```

RESULT 12

```
LOCUS       CR590731                1655 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DE011YH20 of Placenta of Homo sapiens
            (human).
ACCESSION   CR590731
VERSION     CR590731.1  GI:50471538
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1655)
  AUTHORS   Li,W.B., Gruber,C.; Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact: Feng Liang Email: fliang@lifetech.com URL:
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1655)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..1655
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE011YH20"
                     /tissue_type="Placenta"
                     /plasmid="pCMVSPORT_6"
ORIGIN
```

US-10-617-443B-2 (1-334) x CR590731 (1-1655)

Alignment Scores:

```
Pred. No.:                3.83e-114
Score:                    1034.50          Length:        1655
Percent Similarity:       80.18%           Matches:       206
Percent Identity:         80.18%           Conservative:  61
Best Local Similarity:    61.86%           Mismatches:    57
Query Match:              61.61%           Indels:        10
DB:                       3                Gaps:          2
```

```
Qy    4  AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPhe 23
            ||:     |||             :|:||  |||
Db  136  GCCGGAGCCTGCGCCAAGGCAG--GAAGATCCCAACAGTTTGCGCCATAAAATATAACTTT 194

Qy   24  IleAlaAspValValGlyLysIleAlaProAlaValHisGluGluLeuProLeuArg 43
            ||  ||:||||| |||:::||||  |||||||||||          :||
Db  195  ATCCGGAGCAGGTGGTGGAAGAGATCGCCCCCTGCCGTGGGTTCATACGAATTGTTTCGCAAG 254

Qy   44  HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63
            ::|||||          :||      ||:  |||        :|    ||||
Db  255  CTTCCGTTTTCTAAACGAGAGGTGCCGGTGGGTTCTGGGGTTTATTGTGTCGGAA 314

Qy   64  AlaGlyTyrLeuIleLeuThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArg 83
            |||||||||||         :: |||||        :|:||        |:|
Db  315  GATGGACTGATCCTGACAAATGCCCACGTGGTGACAAG 356

Qy   84  GlnGlnLeuLysValGlnLeuGlnAsnGlyAspSerTyrGluAlaThrIleAlaIle 103
            |||||:|||:|||::: ||   ||:|||||        :::|    |||||
Db  357  CACCGGGTCAAAGTTGACCGTGAAGAACGGTTGCCACTTA 416

Qy  104  AspLeuLysSerAspIleAlaThrIleLysAspLysSerProLysLeuProValVal 123
            :::||||||:||||::         |:::|::||   |||
Db  417  GATGAGAAGCAGACATCCGCACTCATCAAAAATTGACCACCAGGGCAAGCTGCCTGTCCTG 476

Qy  124  LeuLeuGlyHisSerAlaAspLeuArgProGlyGluPheValValAlaAlaIleGlySerPro 143
            |||||:|:||:        :::|||||||||         :|||
Db  477  CTGCTTGGCCGCTCCTCAGAGCTCCGGGGCGGAGAGTTCGTCGCCATCGGAAGCCCG 536

Qy  144  PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnArgGluGlyArg 163
            |||::|||||||:|||||        ::|||        |:||:|:|:|||||
Db  537  TTTTCCCTTCAAAAACAGTCACCACCGGGATCGTGAGCACCGCCGCGGCAA 596
```

Qy   30 LysIleAlaProAlaValAlaHisIleLeuGluLeuPheGluArgHisProLeuPheGlyArg 49

Db   62 ... AATGCCCACGTGGTGCACAAC ... (sequence alignment continues)

[The remainder of the page contains a continuation of the query/database (Qy/Db) amino acid and nucleotide sequence alignment blocks with paired residue numbers]

```
Qy    164  GluLeuGlyLeuUvaArgAspSerAspMetAspThrTyrIleGlnThrAspAlaAlaIleIleAsnTyr 183
           |||||::::||||||
Db    597  GAGTCTGGGCGTCCGGAAACTCAGACACATGGACTACACCCAGACCGACGCCATCATCAACTAT 656

Qy    184  GlyAsnSerGlyGlyProLeuUvaAsnLeuAspGlyGluUvaIleIleGlyIleAsnThrLeu 203
           ||||||||||||||||
Db    657  GGAAACTCCGGAGGCCCGTTGAATCTGGACGGTGAAGTAGTGATTGGAATTAACACTTTG 716

Qy    204  LysValThrAlaGlyIleSerPheAlaIleProSerAspArgIleThrArgPheLeuThr 223
           |::||||||||||||
Db    717  AAAGTGACAGCTGGGAATCTCCTTTGCAAATCCCATCTGATAGAATTACAAGGTTCCTCACG 776

Qy    224  GluPheGlnAspIleGlnIleLeuLysAsp------------TrpLysLysArgPheIleGlyIle 240
           ::::::||||                        ::||::||||||||
Db    777  GAGTCCCATGACCGACAGGGCCAAAGGAAAAGCCATCACCAAGAAGAAGTATATTGGTATC 836

Qy    241  ArgMetArgThrIleThrProSerLeuUvaValAspGluLeuLysLysAlaSerAsnProAsnPhe 260
           ::::||||||||::::
Db    837  CGAATGAGTACCACTCAGCCAAAGCCAAGAGCTGAAGGACCGGCACCGGGACTTC 896

Qy    261  ProGluUvaSerGlyIleTyrValGlnGluUvaAlaProAsnSerProSerGlnArg 280
           ::||||||
Db    897  CCAGACGTGATCTCAGGGAGGCTATATAAAATTGAAGTNAATTCCTGATACCCCAGCAGAGCT 956

Qy    281  GlyGlyIleGlnAspGlyAspLysIleIleUvaIleGlyAsnLysAspValAspGluAsnProSer 300
           ::||||||
Db    957  GGTGGTCTCAAGGAAAAACGACGTCATTAATCAACGGACAGTCCGTGGTCTCCGGCC 1016

Qy    301  SerGluLeuGlnUvaValLeuThrGlySerProLeuLeuUvaValArgArgGly 320
           ||||||
Db    1017 AATGATGTCAGGCAGCGTCATTAAAGGAGAAGCACCCTGAACATGGTGGTCTCCGGC 1076

Qy    321  AsnAspAspLeuUvaPheSerIleAlaProGluUvaVal 333
           ::|||
Db    1077 AATGAAGATATCATGATCACAGTGATTCCCGAAGAATT 1115
```

RESULT 13
LOCUS   BC011352
DEFINITION Homo sapiens protease, serine, 11 (IGF binding), mRNA (cDNA clone
            IMAGE:4177882), containing frame-shift errors.
ACCESSION   BC011352
VERSION   BC011352.1 GI:15030191
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 1911)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932

LOCUS   BC011352   1911 bp   mRNA   linear   HTC 19-NOV-2003

REFERENCE   2 (bases 1 to 1911)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 12 Row: h Column: 8
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21327712
            This clone has the following problem: frame shifted.

FEATURES   Location/Qualifiers
     source   1..1911
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4177882"
                     /tissue_type="Brain, anaplastic oligodendroglioma with
                     1p/19q loss"
                     /clone_lib="NCI_CGAP_Brn67"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"

ORIGIN

US-10-617-443B-2 (1-334) x BC011352 (1-1911)

Alignment Scores:
Pred. No.:
Score: 4.70e-114
Percent Similarity: 1034.50
Best Local Similarity: 80.18%
Query Match: 61.86%
DB: 61.61%

Length: 1911
Matches: 206
Conservative: 61
Mismatches: 57
Indels: 10
Gaps: 2

```
Qy    4    AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrLysPheAsnPhe 23
           ||::||
Db    350  GCGGAGCCTGCGGCCAAGGGCAG-GAAGATCCCAACAGTTTGCGCCATAAAATAATAACTTT 408

Qy    24   IleAlaAspValValGluLysIleAlaProAlaLeuUvaAlaHisHisGluLeuPheLeuArg 43
           ||||||
Db    409  ATCGCGGACGTGGTGGAGAAGATCGCCCCTGCCCTGGTTCATATCGAATTGTTTCGCAAG 468

Qy    44   HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheLeuMetSerGlu 63
           ||::||
Db    469  CTTCCGTTTTTCTAAACGAGAGGTGCCGCGTGGGGTTAATTGGTGTCGGAA 528

Qy    64   AlaGlyLeuIleLeuThrAsnAlaHisValValSerSerSerAlaAlaLeuProGlyArg 83
           ||::||
Db    529  GATGACTGATCGTACAAAATGCCCACGTGGTGACCAC------------AAG 570

Qy    84   GlnGlnLeuUvaValGlnLeuUvaLysAsnGlyAspSerTyrGluAlaThrIleLysAspIle 103
           ||::||
Db    571  CACCGGGTCAAAGTTGAGCTGAAGAACACGGTGCCACTTACGAAGCCAAAATCAAGGAGGTG 630

Qy    104  AspLysLysSerAspIleAlaThrLeuLysIleHisProLysLysLeuProValLeu 123
           ||::||||||
```

```
COMMENT        into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by Life Technologies, a
               division of Invitrogen.

REFERENCE      2  (bases 1 to 1931)
AUTHORS        Genoscope
TITLE          Direct Submission
JOURNAL        Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
               BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
               - Web : www.genoscope.cns.fr)
REMARK         1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned

REFERENCE      1  (bases 1 to 1931)
AUTHORS        Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished
REMARK         Contact : Feng Liang Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
               Faraday Avenue

ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
SOURCE         Homo sapiens (human)
KEYWORDS       HTC; CNS1T_cDNA.
VERSION        CR603877.1  GI:50484684
ACCESSION      CR603877
DEFINITION     full-length cDNA clone CS0DI075YN06 of Placenta Cot 25-normalized
               of Homo sapiens (human).
LOCUS          CR603877              1931 bp    mRNA    linear    HTC 21-JUL-2004

RESULT 14
```

```
DEFINITION     full-length cDNA clone CS0DI075YN06 of Placenta Cot 25-normalized
               of Homo sapiens (human).

FEATURES             Location/Qualifiers
     source          1..1931
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI075YN06"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN
```

```
US-10-617-443B-2 (1-334) x CR603877 (1-1931)

Alignment Scores:
Pred. No.:                          1
Score:                              1034.50
Percent Similarity:                 80.18%
Best Local Similarity:              61.86%
Query Match:                        61.61%
DB:                                 3
```

```
Alignment No.:      1
Score:              4.86e-114    Length:       1931
                                 Matches:      206
                                 Conservative: 61
                                 Mismatches:   57
                                 Indels:       10
                                 Gaps:         2
```

Qy   4   AlaLeuProAlaSerAlaGlyLeuHisGlnLeuSerSerProArgTyrArgPheAsnPhe 23
          ::| ::| | | |::| | |||||
Db   454 GCGGAGCCTGCCGCTCCTCAAGGGCAG--GAAGATCCCAACAGTTTGCGCCATAAATATTT 512

Qy   24 IleAlaAspValValGlyLysIleAlaProAlaValAlaHisIleGluLeuPheLeuArg 43
          ||||||||| ||| ::|
Db   513 ATCGCCGACGTGGTGGGAAGAATCGCCCCTGCCGTGGTTCATATCGAAGTGTTTCGCAAG 572

Qy   44 HisProLeuPheGlyArgAsnValProLeuSerSerGlySerGlyPheIleMetSerGlu 63
          | ||||| | :|: | |||||| ||||
Db   573 CTTCCGTTTTCTAAACCAGAGGTGCCCGTGAGTGGGTCTGGGTTTATTGTGTCGGAA 632

Qy   64 AlaGlyLeuIleIleThrAsnAlaHisValAlaSerSerAsnSerAlaAlaProGlyArg 83
          | ||| ::| :|:
Db   633 GATGGACTGATCGTGACAAATGCCCACGTGGTGACCAAC---------------AAG 674

Qy   84 GlnGlnLeuLeuValGlnLeuGlnAsnGlyAspSerThrTyrGlyLeuAlaThrIleLeuLysSerIle 103
          ::||||| |||||| |||:| | ||||||| |||:|:
Db   675 CACCGGCGTCAAAGTTGACGTGAAGAACCGGTGCCACTTACGAAGCCAAAATCAAGGATGTG 734

Qy   104 AspLysSerSerAspIleAlaThrIleLeuHisProLysLysLysLeuProValLeu 123
          ||::|| | ||| :| || :::|
Db   735 GATGAGAAGCAGACATCGCCACTCATCAAAATTGACCACCACCAGGGCAAGCTGCCTGTCCTG 794

Qy   124 LeuLeuGlyHisSerAlaAlaAspLeuArgProGlyGluPheValValAlaAlaIleGlySerPro 143
          ||:|:| |||||| |||:||
Db   795 CTGCTTGGCCGCTCCTCAGAGCTGCGGCCCGGGGAGAGTTCGTGGTGGCCGAAGCCG 854

Qy   144 PheAlaLeuGlnAsnThrValThrThrGlyIleValSerThrAlaGlnAsnArgGluGlyArg 163
          |||| ::| || ||| | :|
Db   855 TTTTCCCTTCAAAACACAGTCACCACCGGGATCGTGAGTACTGCACAGAACCGGGAAGG 914

Qy   164 GluLeuGlyLeuArgAspSerAspMetAspTyrIleGlnThrArgAspAlaIleIleGluAsnTyr 183
          ||||||| ||| :||| ||||||| | |||:|
Db   915 GAGCTGGGGCTCCGCAACTCAGACATGGACTACATCCAGACCCGCGACGCCATCATCGAAAACTAT 974

Qy   184 GlyAsnSerGlyGlyProLeuValAsnLeuAspGlyGluValIleIleGlyIleAsnThrLeu 203
          ||||||| |||||| :| ||| ||||||||
Db   975 GGAAACTCGGGAGGCCCGTTAGTAAACCTGGACGGTGAAGTGATTGGAATTAACACTTTG 1034

Qy   204 LysValThrAlaGlyIleIleSerPheAlaIleProSerArgAspArgIleThrArgPheLeuThr 223
          ||:|||||||| | |||| ||| ||||:|
Db   1035 AAAGTGACAGCTGGAATCCCATCTGATAAAGATAGAATTACAAGGTTCCTCACG 1094

Qy   224 GluPheGlnAspProLysGlnIleIleLysAsp-----------TrpLysLysArgPheIleGlyIle 240
          ||::|| ||| :| |||:|:||||||
Db   1095 GAGTTCCAATGACCCAGACCAAAAGCAAGATATTACAACGTGAAGAAGAGATATTGGTATC 1154

Qy   241 ArgMetArgThrIleThrProSerLeuValAspGluLeuLysAlaSerAsnProAspPhe 260
          ||||| ::| ||||||
Db   1155
```

```
RESULT 15

LOCUS       CR624458               1951 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DK012YA20 of Hela cells Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR624458
VERSION     CR624458.1  GI:50505265
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1951)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1951)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..1951
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK012YA20"
                     /tissue_type="Hela cells Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

US-10-617-443B-2 (1-334) x CR624458 (1-1951)

Alignment Scores:
Pred. No.:
Score:               4.94e-114
Score:               1034.50
Percent Similarity:  80.18%
Best Local Similarity: 61.86%
Query Match:         61.61%
DB:                  3

                Length:      1951
                Matches:     206
                Conservative: 61
                Mismatches:  57
                Indels:      10
                Gaps:        2.
```

```
Qy    1155 CGAATGAGTGTCACTCAGGTCCAGCAAAAGCTGAGGAGGACCGGGGACTTC 1214
                   |||||||||||||||||| ::||||||::|| |||||||||||||||
Db     526 ATCGGCGGACGTGGTGGAGGAGGAGAATCGCGCCCCTCGCGTGGTTCATATCGGAATTGTTTCGGCAAG 585

Qy     261 ProGluValSerSerGlyIleTyrValGlnGluValAlaAlaProAsnSerProSerGlnArg 280
                   |||:::|||:::||   ::|||||||||||::||||||
Db      44 HisProLeuPheGlyArgAsnValProLeuSerSerGlyPheLeuMetSerGlu 63

Qy    1215 CCAGACGTGATCTCAGGAGCGTATATAATTGAAGTAATTCCTGATACCCAGCAGAAGCT 1274
                  ||||||||| ::||||||||||||||||||||||||||||||||||:||
Db     586 CTTCGCGTTTCTAAACGACAGAGAGGTGCCGGTGGCTAGTGTGGGGTTTATTGTGTCGGA 645

Qy     281 GlyGlyIleGlnAsnGlyAspIleLeuValLeuValLysValAsnGlyValArgProLeuValAspSer 300
                   || |::|||| ::|||||:::|||||||||||||||| ::|||
Db      64 AlaGlyValLeuIleLeuIleThrAsnAlaHisValValSerSerAsnSerAlaAlaProGlyArg 83

Qy    1275 GGTGGTCTCAAGGAAAAACGAGCTCATATACTCAGTCGTCCGTGGTCTCCGCC 1334
                   |||||||||:::|||||| ::||| ||||||||||||| ::|||
Db     646 GATGGACTGATCGTGACAAAATGCCCACGTGTGTGACCAAC--------AAG 687

Qy     301 SerGluLeuGlnGluValAlaValLeuThrGluSerProLeuLeuGlnValAlaArgArgGly 320
                   |||||||||| :::|||||||||||||||||| ::|||::|||||||||
Db      84 GlnGlnLeuGluLysValAlaGlnLeuAsnGlyArgProLeuAsnValValAspSer 103

Qy    1335 AATGATGTCAGCGAGGTCATTAAAAGGGAAAAGCACCCTGAACATGGTGTCCGCAGGGGT 1394
                   :::||||||| |:::||||||||||||| ||||||||||||||
Db     104 AspLysLysSerAspLeuAlaThrIleLysIleIleLysHisProLysGlyLysLeuProValLeu 123

Qy     321 AsnAspAlaSerProLeuLeuPheSerIleLeuAlaProGluValVal 333
                   :::||||||||||||||||||||||||
Db    1395 AATGAAGATATCATGATCACAGTGATTCCCGAAGAAATT 1433
```

THIS PAGE BLANK (USPTO)

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:             June 29, 2005, 18:09:44 ; Search time 4078 Seconds
                                              (without alignments)
                                              9688.747 Million cell updates/sec

Title:              US-10-617-443B-1_COPY_1_1038
Perfect Score:      1038
Sequence:           1 cagggactcgaagttcgcag.........ggccctccgccagcgcag 1038
Scoring table:      _____
Searched:           OLIGO NUC
                    Gapop 60.0 , Gapext 60.0
Word Size :         34239544 seqs, 19032134700 residues
                    ___12___

Total number of hits satisfying chosen parameters:    2836061

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :          EST:*
                    1: gb_est1:*
                    2: gb_est2:*
                    3: gb_htc:*
                    4: gb_est3:*
                    5: gb_est4:*
                    6: gb_est5:*
                    7: gb_est6:*
                    8: gb_gss1:*
                    9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                    SUMMARIES

Result          Query
No.     Score   Match   Length  DB  ID              Description
------  -----   -----   ------  --  --              -----------
     1  36      3.5     625     9   AG120009        AG120009 Pan trogl
  c  2  24      2.4     732     5   BQ574834        BQ574834 UI-H-EZI-
  c  3  24      2.3     2534    3   HSM804741       HSM804741 Homo sapi
     4  23      2.2     728     4   BG548488        BG548488 602575077
     5  22      2.2     125     4   BM481069        BM481069 532861 MA
     6  22      2.1     498     6   CD547336        CD547336 B0281B08-
     7  22      2.1     505     6   CD554299        CD554299 B0374A00-
  c  8  22      2.1     744     6   CB528031        CB528031 UI-M-FY0-
  c  9  22      2.1     922     2   BF233030        BF233030 602023960
  c 10  22      2.1     1706    3   BU409267        BU409267 603158028
  c 11  22      2.1     1993    3   AK079109        AK079109 Mus muscu
  c 12  22      2.1     3268    3   BI329065        BI329065 Mus muscu
  c 13  22      2.0     204     4   AK040893        AK040893 Mus muscu
  c 14  22      2.0     207     4   CR516621        CR516621 CR516621
  c 15  21      2.0     238     2   BB200500        BB200500 BB200500
  c 16  21      2.0     243     5   BU756781        BU756781 CH240 137
  c 17  21      2.0     247     9   CC467496        CC467496 CH240 137
  c 18  21      2.0     283     7   AI840198        AI840198 UI-1-CF0-
  c 19  21      2.0     314     2   AW521894        AW521894 UI-M-A00-
  c 20  21      2.0     322     6   AI575250        AI575250 UI-R-B00-
  c 21  21      2.0     322     2   AI705385        AI705385 UI-R-Y0-v
  c 22  21      2.0     322     4   AI705385        AI705385 UI-R-AC1-
  c 23  21      2.0     324     7   CK839023        CK839023 UI-R-AC1-
  c 24  21      2.0     345     7   AW212123        AW212123 uo82f07.x
  c 25  21      2.0     359     4   BI301529        BI301529 UI-R-DN0-
  c 26  21      2.0     375     4   BG381393        BG381393 UI-R-CS0-
  c 27  21      2.0     388     6   CB779415        CB779415 AMGNNUC:U
  c 28  21      2.0     397     2   BF417430        BF417430 UI-R-CN0-
  c 29  21      2.0     405     5   BY367762        BY367762 BY367762
  c 30  21      2.0     407     1   AA967928        AA967928 uh02h11.r
  c 31  21      2.0     426     1   AI553578        AI553578 vw42c02.x
  c 32  21      2.0     428     2   BB860202        BB860202 UI-M-A00-
  c 33  21      2.0     428     7   CN259060        CN259060 170006000
  c 34  21      2.0     470     1   AL134088        AL134088 DKFZp547F
  c 35  21      2.0     508     1   AI415257        AI415257 md17f12.x
  c 36  21      2.0     540     2   BE691054        BE691054 uv55b11.x
  c 37  21      2.0     574     5   BP213364        BP213364 BP213364
  c 38  21      2.0     582     5   BP288294        BP288294 BP288294
  c 39  21      2.0     582     5   BP314507        BP314507 BP314507
  c 40  21      2.0     582     5   BP342411        BP342411 BP342411
  c 41  21      2.0     582     5   BP352419        BP352419 BP352419
  c 42  21      2.0     603     5   BQ210450        BQ210450 UI-R-DYI-
  c 43  21      2.0     604     5   CD371381        CD371381 UI-R-DY0-
  c 44  21      2.0     614     7   CV029025        CV029025 7690 Full
  c 45  21      2.0     619     7   CN259057        CN259057 170004706
```

```
                                           ALIGNMENTS

RESULT 1
AG120009
LOCUS       AG120009                   625 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-128K01.F, genomic survey sequence.
ACCESSION   AG120009
VERSION     AG120009.1  GI:16649174
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 625)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp; URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            clone tracking errors.
            was generated during the R&D process and may have higher chance of
PRIMERS
            Sequencing: -21M13
LIBRARY
            Vector       : pKS145
            R.Site 1     : SacI
            R.Site 2     : SacI
FEATURES             Location/Qualifiers
     source          1..625
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-128K01.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match        3.5%;  Score 36;  DB 9;  Length 625;
Best Local Similarity  100.0%;  Pred. No. 1.2e-06;
```

```
Matches    36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGGGACTCGAAGTTTGCAGTCCTCCACACTCAGTT 36
              ||||||||||||||||||||||||||||||||||||
Db        590 CAGGGACTCGAAGTTTGCAGTCCTCCACACTCAGTT 625
```

RESULT 2
LOCUS   BQ574834/c
DEFINITION   UI-H-EZ1-bbc-f-19-0-UI.s1 NCI_CGAP_Ch2 Homo sapiens cDNA clone
              UI-H-EZ1-bbc-f-19-0-UI 3', mRNA sequence.
ACCESSION   BQ574834
VERSION   BQ574834.1 GI:21478151
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 732)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of
              Orthopaedics
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Clone distribution information can be obtained
              from Dr. M. Bento Soares, bento-soares@uiowa.edu
              The following repetitive elements were found in this cDNA
              sequence: 176-463. >ALU (matched compliment)
              Seq primer: M13 FORWARD
              POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..732
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-EZ1-bbc-f-19-0-UI"
                     /tissue_type="Chondrosarcoma Grade II"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Ch2"
                     /note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Site_1; EcoR I; Site_2: Not I;
                     NCI_CGAP Ch2 is a normalized cDNA library containing the
                     following tissue(s): Chondrosarcoma Grade II. The library
                     was constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_TISSUE=grade-2-chondrosarcoma
                     TAG_LIB=UI-H-EZ1
                     TAG_SEQ=ATCTAATATG"
ORIGIN

```
Query Match    2.3%;  Score 24;  DB 5;  Length 732;
Best Local Similarity  100.0%;
Matches  24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        118 TGAGGGCCACACAGCTAGAAAGCAG 141
              |||||||||||||||||||||||||
Db        493 TGAGGGCCACACAGCTAGAAAGCAG 470
```

RESULT 3
LOCUS   HSM804741   2534 bp   mRNA   linear   HTC 03-AUG-2004
DEFINITION   Homo sapiens mRNA; cDNA DKFZp313M0539 (from clone DKFZp313M0539).
ACCESSION   AL833428
VERSION   AL833428.1 GI:21734070
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2534)
AUTHORS   Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
              Fobo,G., Han,M. and Wiemann,S.
CONSRTM   The German cDNA Consortium
TITLE   Direct Submission
JOURNAL   Submitted (13-JUL-2004) MIPS, Ingolstaedter Landstr.1, D-85764
              Neuherberg, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
              Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
              sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
              consortium of the German Genome Project.
              This clone (DKFZp313M0539) is available at the R2PD Deutsches
              Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
              Please contact RZPD for ordering:
              http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp313M0539
              Further information about the clone and the sequencing project is
              available at http://mips.gsf.de/projects/cdna/.
FEATURES   Location/Qualifiers
     source   1..2534
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313M0539"
     gene   1..2534
                     /gene="DKFZp313M0539"
     CDS   5..>1228
                     /gene="DKFZp313M0539"
                     /codon_start=1
                     /clone_lib="313 (synonym: hlcc2). Vector pTriplEx2; host
                     DH10B; sites SfiIA + SfiIB"
                     /tissue_type="heart"
                     /dev_stage="adult"
                     /note="esophageal cancer associated protein, C-terminus
                     truncated, not fully spliced"
                     /product="hypothetical protein"
                     /protein_id="CAH10399.1"
                     /db_xref="GI:5094538"
                     /translation="MAVFWHSRNRNYKAEFASCRLEAVPLEFGDYHPLKPITVTESK
                     TKKVNRKGSTSTSSSSSSSVVDPLSSVLDGTGTDPLSMFAATADPPALLAAMDSSRKR
                     DRDDNSVVGSDFEPWTNKRGEILARYTTEKLSINLFMGSEKGKAGTATLAMSEKVRT
                     RLEELDSEBEGSQKELLNLTQQDYVNRIEELNQSLKDAWASDQKVRALKIVIQCSKLL
                     SDTSVIQFYPSKFVLLITDILDTFGKLVYERLFSMCVDSRSVLPDHFSPENANDTAKET
                     CLNWFKIASIRELIPRFVEASILKCNKFLSKTGISECLPRLTCMIRGIGDPLVSVY
                     ARAYLCRVGMEVAPHLKETLNKNFDFLLTFRQIHGDTVQNQLVVQGVELPSYLPLYP
                     PAMDWIFQCISYHAPE"
ORIGIN

```
Query Match    2.3%;  Score 24;  DB 3;  Length 2534;
Best Local Similarity  100.0%;
Matches  24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        118 TGAGGGCCACACAGCTAGAAAGCAG 141
              |||||||||||||||||||||||||
Db       1468 TGAGGGCCACACAGCTAGAAAGCAG 1491
```

RESULT 4
LOCUS   BG548488   728 bp   mRNA   linear   EST 04-APR-2001
```
Query Match    2.3%;  Score 24;  DB 3;  Length 2534;
```

DEFINITION   6025750777F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703322 5',
              mRNA sequence.
ACCESSION   BG548488
VERSION   BG548488.1   GI:13547153
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 728)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: CLONTECH Laboratories, Inc.
              cDNA Library Preparation: CLONTECH Laboratories, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1541 row: i column: 19
              High quality sequence stop: 695.
FEATURES   Location/Qualifiers
     source   1..728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4703322"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
                     SfiI (ggcgcgcctcggcc); Site 2: SfiI (ggccatatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
                     5'-ATTCTAGAGGCGCGAGGCGGCCGACACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN
Qy   139 CAGCCAGGGCCCAGCCGAACCCT 161
                   |||||||||||||||||||||||
Db   491 CAGCCAGGGCCCAGCCGAACCCT 513

RESULT 5
LOCUS   BM481069   125 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION 532861 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BM481069
VERSION   BM481069.1   GI:18531397
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1 (bases 1 to 125)
AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
TITLE   Sequence evaluation and construction of a gene index for cattle

Query Match   2.2%;   Score 23;   DB 4;   Length 728;
Best Local Similarity   100.0%;   Pred. No. 5.4;
Matches 23;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

JOURNAL   Genome Res. 11 (4), 626-630 (2001)
MEDLINE   21180013
PUBMED   11282978
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 2 row: G column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
     source   1..125
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4BOV"
                     /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
ORIGIN
Qy   329 GGTCCTGGGCCAGCCCACAGCCCCCT 350
                 ||||||||||||||||||||||||||
Db   6   GGTCCTGGGCCAGCCCACAGCCCCCT 27

RESULT 6
LOCUS   CD547336   498 bp   mRNA   linear   EST 11-JUN-2003
DEFINITION BC281B08-5 NIA Mouse Embryonic Germ Cell cDNA Library (Long) Mus
            musculus cDNA clone NIA:B028IB08 IMAGE:30111187 5', mRNA sequence.
ACCESSION   CD547336
VERSION   CD547336.1   GI:31595071
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 498)
AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE   Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE   21429098
PUBMED   11544199
COMMENT   Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0281 row: B column: 08
            Seq primer: M13 Reverse
            High quality sequence stop: 498
            POLYA=No.
FEATURES   Location/Qualifiers
     source   1..498
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"

Query Match   2.1%;   Score 22;   DB 4;   Length 125;
Best Local Similarity   100.0%;   Pred. No. 18;
Matches 22;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Left column

```
          /db_xref="niaEST:B0281808-5"
          /db_xref="taxon:10090"
          /clone_lib="NIA:B0281808 IMAGE:30111187"
          /sex="male"
          /dev_stage="embryonic day 8"
          /lab_host="DH10B"
          /clone_lib="NIA Mouse Embryonic Germ Cell cDNA Library
          (Long)"
          /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
          Site 2: NotI; Mouse cDNA project by the Laboratory of
          Genetics, National Institute on Aging (NIA), Intramural
          Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
          This is a long-transcript enriched cDNA library [Ref.
          Genome Res. 11: 1553-1558 (2001). [PMID: 11541991]). Total
          RNAs were obtained from Dr.Mark G. Carter (NIH/NIA-IRP).
          EG cells were cultured at 37 C, 5% CO2 in DMEM
          supplemented with 15% ES cell-qualified FBS, 0.1mM
          non-essential amino acids, 2 mM glutamine,
          penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM
          beta-mercaptoethanol, and 1000000 units of LIF per liter.
          Double-stranded cDNAs were synthesized with an Oligo(dT)
          primer [Invitrogen].
          5'-pGACTAGTTCTAGATCGCGAGCGGCCGCGGGCGCCCCTTTTTTTTTTTTTTTTT-3'] from
          2.5 ug of total RNA, treated with T4 DNA polymerase, and
          purified by ethanol-precipitation. The cDNAs were ligated
          to Lone-linker Li-Sal4, purified by phenol/chloroform, and
          separated from free linkers by Centricon 100. Then, the
          cDNAs were amplified by long-range high fidelity PCR using
          Ex Taq polymerase (Takara) with a primer Sal4-S. The
          products were purified by phenol/chloroform and Centricon
          100. The cDNAs were digested with SalI and NotI enzymes
          and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
          vector. The DH10B E. coli host was transformed with the
          ligation mixture by the standard chemical method. The
          average insert size is about 4.0 kb. The library was
          constructed by Yulan Piao."

ORIGIN
```

```
RESULT 7
LOCUS      CD554299/c

Query Match     2.1%;   Score 22;   DB 6;   Length 498;
Best Local Similarity  100.0%;   Pred. No. 17;
Matches 22;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   65 AGTCCCATTTTTCAGATGAGGA  86
        ||||||||||||||||||||||
Db  462 AGTCCCATTTTTCAGATGAGGA 441
```

```
LOCUS      CD554299     505 bp   mRNA   linear   EST 11-JUN-2003
DEFINITION B0374A09-5 NIA Mouse E9.5 Whole Embryo cDNA Library
           (Long) Mus musculus cDNA clone NIA:B0374A09
           IMAGE:30436136 5', mRNA sequence.
ACCESSION  CD554299
VERSION    CD554299.1  GI:3160030
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 505)
AUTHORS    Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE      Construction of long-transcript enriched cDNA libraries from
           submicrogram amounts of total RNAs by a universal PCR amplification
           method
JOURNAL    Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE    21429098
PUBMED     11554199
COMMENT    Contact: Dawood B. Dudekula
           Laboratory of Genetics
           National Institute on Aging/National Institutes of Health
           333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
           Email: cdna@lgsun.grc.nia.nih.gov
           Plate: B0374 row: A column: 09
```

Right column

```
FEATURES             Location/Qualifiers
     source          1..505
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0374A09-5"
                     /db_xref="niaEST:B0374A09-5"
                     /clone_lib="NIA:B0374A09 IMAGE:30436136"
                     /tissue_type="E9.5 whole embryo"
                     /dev_stage="E9.5 whole embryo including extraembryonic tissues
                     at 9.5-days postcoitum"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse E9.5 Whole Embryo cDNA Library
                     (Long)"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
                     Site 2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library [Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11541991]). Total
                     RNAs were extracted from a pool of 16 embryos at 9.5-days
                     postcoitum. Double-stranded cDNAs were synthesized with an
                     Oligo(dT) primer [Invitrogen].
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCGGGCGCCCCTTTTTTTTTTTTTTTTT-3']
                     from 1 ug of total RNA, treated with T4 DNA polymerase,
                     and purified by ethanol-precipitation. The cDNAs were
                     ligated to Lone-linker Li-Sal4, purified by
                     phenol/chloroform, and separated from free linkers by
                     Centricon 100. Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S. The products were purified
                     by phenol/chloroform and Centricon 100. The cDNAs were
                     digested with SalI and NotI enzymes and cloned into
                     SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E.
                     coli host was transformed with the ligation mixture by the
                     standard chemical method. The average insert size is about
                     3.0kb. The library was constructed by Yulan Piao."

Seq primer: M13 Reverse
High quality sequence stop: 505
POLYA=No.

ORIGIN
```

```
RESULT 8
LOCUS      CB528031/c

Query Match     2.1%;   Score 22;   DB 6;   Length 744;
Best Local Similarity  100.0%;   Pred. No. 17;
Matches 22;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   65 AGTCCCATTTTTCAGATGAGGA  86
        ||||||||||||||||||||||
Db  462 AGTCCCATTTTTCAGATGAGGA 441
```

```
LOCUS      CB528031     744 bp   mRNA   linear   EST 09-JUL-2003
DEFINITION UI-M-FY0-cfk-p-20-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA
           clone IMAGE:6850917 5', mRNA sequence.
ACCESSION  CB528031
VERSION    CB528031.1  GI:29361504
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 744)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Dr. Jim Lin, University of Iowa
           cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
           cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
           DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
```

FEATURES
    source   Location/Qualifiers
                    1..744
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6"
                    /clone="whole brain"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:6850917"
                    /tissue_type="whole brain"
                    /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                    /clone_lib="NIH_BMAP_FYO"
                    /lab_host="DH10B (T1 phage resistant)"
                    /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                    Site_2: Not I; The library was constructed according
                    Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                    1996. Denatured RNA was size fractionated on a 1% agarose
                    gel. First strand cDNA synthesis was primed with oligo-dT
                    primer containing a Not I site. Double strand cDNA was
                    size selected according to mRNA size fraction, ligated
                    with EcoR I adaptor, digested with NotI and then cloned
                    directionally into pYX-Asc vector. The library tag
                    sequence located between the Not I site and the polyA tail
                    is AGCGAGACAG. This library was created for the University
                    Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the
                    Developing Mouse Nervous System', supported by National
                    Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                    program coordinator."

Seq primer: pYX-5.

Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousef1.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

ORIGIN

Qy   65 AGTCCCATTTTTCAGATGAGGA 86
            ||||||||||||||||||||||
Db   346 AGTCCCATTTTTCAGATGAGGA 325

Matches 22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
Best Local Similarity   100.0%; Pred. No. 17;
Query Match   2.1%;   Score 22;   DB 6;   Length 744;

ORIGIN

COMMENT   High quality sequence stop: 622.
            Plate: LLAM9436 row: h column: 15
            http://image.llnl.gov
            found through the I.M.A.G.E. Consortium/LLNL at:
            Clone distribution: MGC clone distribution information can be
            DNA Sequencing by: Incyte Genomics, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Library Preparation: Life Technologies, Inc.
            Tissue Procurement: Jeffrey B. Green, M.D.
            Email: cgapbs-r@mail.nih.gov
            Contact: Robert Strausberg, Ph.D.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   Unpublished (1999)
TITLE   Generation and initial analysis
AUTHORS
REFERENCE   1 (bases 1 to 922)
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM   Mus musculus
SOURCE   Mus musculus (house mouse)
KEYWORDS   EST.
VERSION   BP233030.1 GI:1113008
ACCESSION   BP233030
            mRNA sequence.
DEFINITION 602023960F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4159166 5',
LOCUS   BP233030   922 bp   mRNA   linear   EST 14-NOV-2000
BF233030/c

RESULT 9

Matches 22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
Best Local Similarity   100.0%; Pred. No. 17;
Query Match   2.1%;   Score 22;   DB 2;   Length 922;

ORIGIN

Qy   65 AGTCCCATTTTTCAGATGAGGA 86
            ||||||||||||||||||||||
Db   560 AGTCCCATTTTTCAGATGAGGA 539

RESULT 10
BF233030

FEATURES
    source   Location/Qualifiers
                1..922
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4159166"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_lib="NCI_CGAP_Li9"
                /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.9 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP library."

ORIGIN

COMMENT   Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 0161208930
            Fax: 0161236040
            Email: Simon.Hubbard@umist.ac.uk.
PUBMED   12455392
MEDLINE   22235534
JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
TITLE   A Comprehensive Collection of Chicken cDNAs
AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
REFERENCE   1 (bases 1 to 1706)
            Phasianidae; Gallus.
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM   Gallus gallus
SOURCE   Gallus gallus (chicken)
KEYWORDS   EST.
VERSION   BU409267.1 GI:25901938
ACCESSION   BU409267
            sequence.
DEFINITION 603158028F1 CSEQRBL03 Gallus gallus cDNA clone ChEST175g8 5', mRNA
LOCUS   BU409267   1706 bp   mRNA   linear   EST 29-NOV-2002
BU409267

RESULT 10

FEATURES
    source   Location/Qualifiers
                1..1706
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="Layer and broiler"
                /db_xref="taxon:9031"
                /clone="ChEST175g8"
                /sex="Male and female"
                /tissue_type="Chondrocytes isolated from growth plate
                cartilage"
                /dev_stage="adult"
                /lab_host="DH10B"
                /clone_lib="CSEQRBL03"
                /note="Vector: pBluescript II KS(+); Site_1: EcoRI;
                Site_2: NotI; Modification of pBluescript II KS(+)
                [Stratagene] vector to accommodate cDNA produced with the
                T-trimmed protocol (Construction of uni-directionally
                cloned cDNA libraries from messenger RNA for improved 3'
                end DNA sequencing by Glenn Fu, et al. U.S. Patent #
                6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
                Ligate in double stranded adaptor containing BsgI and
                BamHI sites [5'ggccgcgtgcagcccccggatccgaaaaaag]
                [5'aattccttctttcggatccgggggctgcacgc]"

ORIGIN

RESULT 11

Query Match
Best Local Similarity   2.1%; Score 22; DB 5; Length 1706;
Matches 22; Conservative 0; Mismatches   0; Indels   0; Gaps   0;

Qy   161 TGGTGTGTGCAGCCCCCAGCCC 182
            |||||||||||||||||||||
Db   255 TGGTGTGTGCAGCCCCCAGCCC 276

LOCUS   AK079109   1993 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region
            and neck cDNA, RIKEN full-length enriched library, clone:9430008C03
            product:unknown EST, full insert sequence.
ACCESSION   AK079109
VERSION   AK079109.1 GI:26098280
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 1993)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,Y., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

  TITLE   Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in Riken.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
     source   1..1993
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9430008C03"
                     /tissue_type="embryonic body between diaphragm region and
                     neck"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     misc_feature   1..1993
                     /note="unknown EST (GB|BI251132, evidence: BLASTN, 98%,
                     match=822)"
ORIGIN

RESULT 12

Query Match
Best Local Similarity   2.1%; Score 22; DB 3; Length 1993;
Matches 22; Conservative 0; Mismatches   0; Indels   0; Gaps   0;

Qy   65 AGTCCCATTTTTCAGATGAGGA 86
            |||||||||||||||||||||
Db   634 AGTCCCATTTTTCAGATGAGGA 613

LOCUS   AK040893   3268 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus adult male aorta and vein cDNA, RIKEN full-length
            enriched library, clone:A530040I21 product:unknown EST, full insert
            sequence.
ACCESSION   AK040893
VERSION   AK040893.1 GI:26088138
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

AUTHORS   Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system -384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The PANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3268)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kawaguchi,J., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Nunazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group; RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in Riken
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
     source 1. .3268
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="A530040121"
             /sex="male"
             /tissue_type="aorta and vein"
             /clone_lib="RIKEN full-length enriched mouse cDNA library"
             /dev_stage="adult"
     misc_feature 1. .3268
             /note="unknown EST (GB|BI251132, evidence: BLASTN, 98%,
             match=822)"
ORIGIN

Query Match   2.1%;   Score 22;   DB 3;   Length 3268;
Best Local Similarity 100.0%;   Pred. No. 17;
Matches 22;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   65 AGTCCCATTTTTCAGATGAGGA   86
         ||||||||||||||||||||||
Db 625 AGTCCCATTTTTCAGATGAGGA 604

RESULT 13
BI329065/c
BI329065   602980483F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5133309 5',
mRNA sequence.
LOCUS   BI329065   204 bp   mRNA   linear   EST 30-JUL-2001
DEFINITION BI329065 602980483F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5133309 5',
            mRNA sequence.
ACCESSION   BI329065
VERSION   BI329065.1 GI:15013722
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 204)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11327 row: e column: 204.
            High quality sequence stop: 204.
FEATURES   Location/Qualifiers
     source 1. .204
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:5133309"
             /clone_lib="NCI_CGAP_Li9"
             /lab_host="DH10B (T1 phage-resistant)"
             /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
             Average insert size 1.9 kb. Constructed by Life
             Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Query Match   2.0%;   Score 21;   DB 4;   Length 204;
Best Local Similarity 100.0%;   Pred. No. 57;
Matches 21;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy 903 TGTGGGCTGAGGCTGGCTCTG 923
         |||||||||||||||||||||
Db   48 TGTGGGCTGAGGCTGGCTCTG   28

RESULT 14
CR516621
LOCUS   CR516621   207 bp   mRNA   linear   EST 07-JUL-2004
DEFINITION CR516621 Mouse pBluescript Lion Mus musculus cDNA clone
            LIONp462B0815 3', mRNA sequence.
ACCESSION   CR516621
VERSION   CR516621.1 GI:49908681
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 207)
AUTHORS   Henrich,J., Hermans,J., Kranz,H., Loebbert,R., Schlueter,T.,
            Schuette,D., Weindel,M., Heil,O., Ebert,L., Neubert,P., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
TITLE   Mouse ArrayTAG cDNA
JOURNAL   Unpublished (2004)
COMMENT   Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany
Email: www.rzpd.de
RZPD; LIONp462B0815.
RZPDLIB;
Mouse ArrayTAG cDNA
http://www.rzpd.de/cgi-
bin/products/showLib.pl.cgi/response?libNo=462 Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 3269 100
Fax: +49 30 3269 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
RP: CAGGAAACAGCTATGAC.

FEATURES
source   Location/Qualifiers
                     1..207
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="LIONp462B0815"
                     /lab_host="DH10B"
                     /clone_lib="Mouse pBluescript Lion"

ORIGIN

Query Match   2.0%; Score 21; DB 7; Length 207;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   903 TGTGGGGCTGAGGCTGGCTCTG 923
          ||||||||||||||||||||||
Db   140 TGTGGGGCTGAGGCTGGCTCTG 160

RESULT 15
LOCUS   BB200500   238 bp   mRNA   linear   EST 30-JUN-2000
DEFINITION BB200500 RIKEN full-length enriched, 0 day neonate thymus Mus
            musculus cDNA clone A430025A01 3' similar to AF172275 Mus musculus
            FUS2 (Fus2) mRNA, mRNA sequence.
ACCESSION   BB200500
VERSION   BB200500.1 GI:8865453
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 238)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
            Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
            Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
            Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
            Shibata,K., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Sugahara,Y., Suzuki,H., Tagawa,A.,
            Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
            Watanabe,S., Yamanaka,I., Yamamura,T., Yano,R., Yasunishi,A.,
            Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source   Location/Qualifiers
                     1..238
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="A430025A01"
                     /tissue_type="thymus"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     thymus"
                     /note="Site 1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer (5'
                     GAGAGAGAAGGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'), cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence (5' GAGGAGGATTCTCGACTTAATTAARTTAATCCCCCCCCCCCC
                     3'). cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."

ORIGIN

Query Match   2.0%; Score 21; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   903 TGTGGGGCTGAGGCTGGCTCTG 923
          ||||||||||||||||||||||
Db   100 TGTGGGGCTGAGGCTGGCTCTG 80

Search completed: June 29, 2005, 21:07:12
Db
Job time : 4086 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 18:13:24 ; Search time 219 Seconds
                 (without alignments)
                 7755.503 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 cagggactcgaagttgcag.........ggcccttcccgccagcgcag 1038
Scoring table: OLIGO NUC
                 Gapop 60.0 , Gapext 60.0
Searched: 1202784 seqs, 818138359 residues
Word size: 12

Total number of hits satisfying chosen parameters: 83567

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :   Issued Patents NA:*
                 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
c 1	22	2.1	116652	4	US-09-949-016-13413	4	Sequence 13413, A
c 2	21	2.0	65744	4	US-09-949-016-12591	4	Sequence 12591, A
c 3	21	2.0	65745	4	US-09-949-016-15871	4	Sequence 15871, A
4	21	2.0	818128	4	US-09-866-108A-15742	4	Sequence 15742, A
5	21	2.0	818128	4	US-09-949-016-15546	4	Sequence 15546, A
6	21	2.0	818128	4	US-09-949-016-14547	4	Sequence 14547, A
7	21	2.0	818128	4	US-09-949-016-14548	4	Sequence 14548, A
8	21	2.0	818128	4	US-09-949-016-14549	4	Sequence 14549, A
9	21	2.0	818128	4	US-09-949-016-14550	4	Sequence 14550, A
10	21	2.0	818128	4	US-09-949-016-14551	4	Sequence 14551, A
11	21	2.0	818128	4	US-09-949-016-14552	4	Sequence 14552, A
12	21	2.0	818128	4	US-09-949-016-14553	4	Sequence 14553, A
13	21	2.0	818128	4	US-09-949-016-14554	4	Sequence 14554, A
14	21	2.0	818128	4	US-09-949-016-14555	4	Sequence 14555, A
15	21	2.0	818128	4	US-09-949-016-14556	4	Sequence 14556, A
16	21	2.0	818128	4	US-09-949-016-14557	4	Sequence 14557, A
17	21	2.0	818128	4	US-09-949-016-14558	4	Sequence 14558, A
18	21	2.0	818128	4	US-09-949-016-14559	4	Sequence 14559, A
19	21	2.0	818128	4	US-09-949-016-14560	4	Sequence 14560, A
20	21	2.0	818128	4	US-09-949-016-14561	4	Sequence 14561, A
21	21	2.0	818128	4	US-09-949-016-14562	4	Sequence 14562, A
22	21	2.0	818128	4	US-09-949-016-14565	4	Sequence 14565, A
23	21	2.0	818128	4	US-09-949-016-14566	4	Sequence 14566, A
24	21	2.0	818128	4	US-09-949-016-14567	4	Sequence 14567, A
c 25	20	1.9	601	4	US-09-949-016-46836	4	Sequence 46836, A
c 26	20	1.9	601	4	US-09-949-016-58443	4	Sequence 58443, A
c 27	20	1.9	17353	4	US-09-949-016-13437	4	Sequence 13437, A
28	20	1.9	28555	4	US-09-949-016-13106	4	Sequence 13106, A
29	19	1.8	89	4	US-09-513-999C-16411	4	Sequence 16411, A
30	19	1.8	500	4	US-09-866-108A-15742	4	Sequence 15742, A
31	19	1.8	558	1	US-08-244-492A-4	4	Sequence 4, Appli
32	19	1.8	558	1	US-08-709-915-5	4	Sequence 5, Appli
c 33	19	1.8	601	4	US-09-949-016-26845	4	Sequence 26845, A
34	19	1.8	601	4	US-09-949-016-47781	4	Sequence 47781, A
c 35	19	1.8	601	4	US-09-949-016-55840	4	Sequence 55840, A
c 36	19	1.8	601	4	US-09-949-016-56305	4	Sequence 56305, A
37	19	1.8	601	4	US-09-949-016-155557	4	Sequence 155557,
38	19	1.8	601	4	US-09-949-016-155656	4	Sequence 155656,
39	19	1.8	637	4	US-09-517-431E-18	4	Sequence 18, Appl
40	19	1.8	1215	3	US-09-091-952A-29	4	Sequence 29, Appl
c 41	19	1.8	25202	4	US-09-949-016-13151	4	Sequence 13151, A
c 42	19	1.8	27380	4	US-09-949-016-12877	4	Sequence 12877, A
c 43	19	1.8	27383	4	US-09-949-016-14393	4	Sequence 14393, A
c 44	19	1.8	38078	4	US-09-949-016-12429	4	Sequence 12429, A
c 45	19	1.8	38084	4	US-09-949-016-16459	4	Sequence 16459, A

ALIGNMENTS

RESULT 1
US-09-949-016-13413/c
; Sequence 13413, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13413
; LENGTH: 116652
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(116652)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13413

Query Match   2.1%; Score 22; DB 4; Length 116652;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches   22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   931 ACAGGGGGTCTCAGGAAGAGCCT 952
               |||||||||||||||||||||||
Db   75459 ACAGGGGGTCTCAGGAAGAGCCT 75438

RESULT 2
US-09-949-016-12591
; Sequence 12591, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12591
; LENGTH: 65744
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(65744)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12591

Qy        124 CACACAGCTTAGAAAAGCAGCCA 144
              |||||||||||||||||||||
Db      44923 CACACAGCTTAGAAAAGCAGCCA 44943

Query Match      100.0%;  Score 21;  DB 4;  Length 65744;
Best Local Similarity   100.0%;  Pred. No. 2.9;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 3
US-09-949-016-15871
; Sequence 15871, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15871
; LENGTH: 65745
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(65745)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15871

Qy        124 CACACAGCTTAGAAAAGCAGCCA 144
              |||||||||||||||||||||
Db      44923 CACACAGCTTAGAAAAGCAGCCA 44943

Query Match      100.0%;  Score 21;  DB 4;  Length 65745;
Best Local Similarity   100.0%;  Pred. No. 2.9;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 4
US-09-949-016-14546
; Sequence 14546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14546
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14546

Qy        120 AGGCCCACACAGCTTAGAAAAGCA 140
              ||||||||||||||||||||||
Db     116875 AGGCCCACACAGCTTAGAAAAGCA 116895

Query Match      100.0%;  Score 21;  DB 4;  Length 818128;
Best Local Similarity   100.0%;  Pred. No. 2.9;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 5
US-09-949-016-14547
; Sequence 14547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14547
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14547

Qy        120 AGGCCCACACAGCTTAGAAAAGCA 140
              ||||||||||||||||||||||
Db     116875 AGGCCCACACAGCTTAGAAAAGCA 116895

Query Match      100.0%;  Score 21;  DB 4;  Length 818128;
Best Local Similarity   100.0%;  Pred. No. 2.9;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 6
US-09-949-016-14548
; Sequence 14548, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

RESULT 7
US-09-949-016-14549
; Sequence 14549, Application US/0949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 14549
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549

Query Match   2.0%;   Score 21;   DB 4;   Length 818128;
Best Local Similarity 100.0%;   Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   120 AGGCCACACAGCTAGAAAAGCA 140
                |||||||||||||||||||||
Db   116875 AGGCCACACAGCTAGAAAAGCA 116895

RESULT 8
US-09-949-016-14550
; Sequence 14550, Application US/0949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 14550
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550

Query Match   2.0%;   Score 21;   DB 4;   Length 818128;
Best Local Similarity 100.0%;   Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   120 AGGCCACACAGCTAGAAAAGCA 140
                |||||||||||||||||||||
Db   116875 AGGCCACACAGCTAGAAAAGCA 116895

RESULT 9
US-09-949-016-14551
; Sequence 14551, Application US/0949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 14551
;   LENGTH: 818128
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(818128)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14551

Query Match   2.0%;   Score 21;   DB 4;   Length 818128;
Best Local Similarity 100.0%;   Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   120 AGGCCACACAGCTAGAAAAGCA 140
                |||||||||||||||||||||
Db   116875 AGGCCACACAGCTAGAAAAGCA 116895

RESULT 10
US-09-949-016-14552
; Sequence 14552, Application US/0994901b
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 14552
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552

Query Match   2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity   100.0%; Pred. No. 2.9;
  Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   120 AGGCCACACAGCTAGAAAGCA 140
                ||||||||||||||||||||||
Db   116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 11
US-09-949-016-14553
; Sequence 14553, Application US/0994901b
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 14553
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553

Query Match   2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity   100.0%; Pred. No. 2.9;
  Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   120 AGGCCACACAGCTAGAAAGCA 140
                ||||||||||||||||||||||
Db   116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 12
US-09-949-016-14554
; Sequence 14554, Application US/0994901b
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 14554
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554

Query Match   2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity   100.0%; Pred. No. 2.9;
  Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   120 AGGCCACACAGCTAGAAAGCA 140
                ||||||||||||||||||||||
Db   116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 13
US-09-949-016-14555
; Sequence 14555, Application US/0994901b
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 14555
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555

Query Match   2.0%; Score 21; DB 4; Length 818128;
Best Local Similarity   100.0%; Pred. No. 2.9;
  Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   120 AGGCCACACAGCTAGAAAGCA 140
                ||||||||||||||||||||||

```
Qy        120 AGGCCACACAGCTAGAAAGCA 140
              ||||||||||||||||||||
Db     116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 14
US-09-949-016-14556
; Sequence 14556, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14556
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556

Qy        120 AGGCCACACAGCTAGAAAGCA 140
              ||||||||||||||||||||
Db     116875 AGGCCACACAGCTAGAAAGCA 116895

Query Match          2.0%;  Score 21;  DB 4;  Length 818128;
Best Local Similarity  100.0%;  Pred. No. 2.9;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        120 AGGCCACACAGCTAGAAAGCA 140
              ||||||||||||||||||||
Db     116875 AGGCCACACAGCTAGAAAGCA 116895

RESULT 15
US-09-949-016-14557
; Sequence 14557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14557
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14557
```

```
Query Match          2.0%;  Score 21;  DB 4;  Length 818128;
Best Local Similarity  100.0%;  Pred. No. 2.9;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        120 AGGCCACACAGCTAGAAAGCA 140
              ||||||||||||||||||||
Db     116875 AGGCCACACAGCTAGAAAGCA 116895

Search completed: June 29, 2005, 21:11:05
Job time : 234 secs
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6.
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 15:40:04 ; Search time 675 Seconds
               (without alignments)
               9103.247 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 caggggactcgaagttgcag ....ggcccttccgccagccgcag 1038
Scoring table: Gapop 60.0, Gapext 60.0
Searched: 4390206 seqs, 295987.0667 residues

Database :   N_Geneseq_16Dec04:
               1: geneseqn1980s:*
               2: geneseqn1990s:*
               3: geneseqn200s:*
               4: geneseqn2001a:*
               5: geneseqn2001bs:*
               6: geneseqn2001as:*
               7: geneseqn2002as:*
               8: geneseqn2002bs:*
               9: geneseqn2003as:*
              10: geneseqn2003bs:*
              11: geneseqn2003cs:*
              12: geneseqn2004as:*
              13: geneseqn2004bs:*

Post-processing: Listing first 45 summaries
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Total number of hits satisfying chosen parameters: 436444

Word size: 12

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%	Query Match Length	DB	ID	Description
1	1038	100.0	3006	12	ADJ11355	Adj11355 Human PRS
2	22	2.1	1379	3	AAF21695	Aaf21695 Human bre
3	22	2.1	458	12	ADD47518	Add47518 Human gen
4	22	2.1	458	10	ADD47514	Add47514 Human gen
5	22	2.1	112453	13	ABD32665	Abd32665 Human can
6	21	2.0	34	12	ADJ11359	Adj11359 PCR prime
7	21	2.0	244	12	ADG99925	Adg99925 Kidney di
8	21	2.0	1795	6	ABL99890	Abl99890 Human sec
9	21	2.0	1795	6	ADQ87121	Adq87121 Human tum
10	21	2.0	3157	12	ABZ36074	Abz36074 Human sec
11	21	2.0	3983	8	ADM87103	Adm87103 Human pro
12	21	2.0	4030	12	ADQ80254	Adq80254 Hermansky
13	21	2.0	127145	13	ACF62741	Acf62741 Cancer ba
14	21	2.0	189013	8	ADB20856	Adb20856 MRP1 base
15	21	2.0	189013	10	ADB87945	Adb87945 Human UGT
16	21	2.0	189013	10	ADB96928	Adb96928 Human MDR
17	21	2.0	189013	10	ADB92119	Adb92119 Human MDR
18	21	1.9	20	12	ADJ11361	Adj11361 PCR prime
19	20	1.9	301	4	AAH57233	Aah57233 Human int
20	20	1.9	836	2	AAQ06484	Aaq06484 Intestina
c 21	20	1.9	836	2	AAQ12218	Aaq12218 SMJC-41 1
c 22	20	1.9	3294	4	AAH57441	Aah57441 Human int
c 23	20	1.9	15720	4	AAI29504	Aai29504 C899P det
c 24	20	1.9	15720	6	ABl62358	Abl62358 Colon ade
c 25	20	1.9	15720	8	ABZ33690	Abz33690 Human col
c 26	20	1.9	15720	8	ACD13431	Acd13431 Human DNA
c 27	20	1.9	15720	10	AAD63141	Aad63141 Human muc
c 28	20	1.9	15720	10	AAD62742	Aad62742 Human muc
c 29	20	1.9	15720	11	ADP88237	Adp88237 Colon can
c 30	20	1.9	15720	11	ADP88236	Adp88236 Colon can
c 31	20	1.9	15720	11	ADQ29628	Adq29628 Human col
c 32	20	1.9	15720	12	ADQ80264	Adq80264 Intestina
c 33	20	1.9	79860	13	ADR86701	Adr86701 Human Eph
c 34	20	1.9	26000	13	ADR82648	Adr82648 Human Eph
c 35	20	1.9	55875	12	ADQ59503	Adq59503 Human can
c 36	20	1.9	87761	11	ACN44122	Acn44122 Human gen
c 37	20	1.9	94531	11	ADL27164	Adl27164 Human gen
c 38	20	1.9	94529	9	ADA03086	Ada03086 Human hCG
c 39	20	1.9	94529	10	ADB72824	Adb72824 Human hCG
c 40	20	1.9	94529	9	ADA66370	Ada66370 Human hCG
c 41	20	1.9	94529	8	AAD54480	Aad54480 Human hCG
c 42	19	1.9	117962	8	AAC12336	Aac12336 Human sec
c 43	19	1.8	89	12	ADM72179	Adm72179 Human NTR
c 44	19	1.8	251	12	ADS65486	Ads65486 Corn seed
c 45	19	1.8	290	6	ABV95882	Abv95882 Human pan

ALIGNMENTS

RESULT 1
ADJ11355

ID   ADJ11355 standard; cDNA; 3006 BP.
XX
AC   ADJ11355;
XX
DT   20-MAY-2004 (first entry)
XX
DE   Human PRSS11-L cDNA encoding a serine protease SeqID 1.
XX
KW   human; PRSS11-L; gene; ss; S2 serine protease; S2/HtrA;
KW   cellular physiology; apoptosis; gene therapy; osteopathic; stress;
KW   apoptotic; osteoarthritic; antiarthritic; tranquiliser.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   CDS   1011..2015
XX   /*tag=   a
XX   /product= "PRSS11-L protein"
XX
PN   US2004005659-A1.
XX
PD   08-JAN-2004.
XX
PF   03-JUL-2002; 2002US-0189099.
XX
PR   03-JUL-2002; 2002US-0189099.
XX
PA   (DARR/) DARROW A L.
PA   (QLJ/) QL J.
PA   (CHEN/) CHEN C.
PA   (ANDR/) ANDRADE-GORDON P.
XX
PI   Darrow AL, Ql J, Chen C, Andrade-Gordon P;
XX
DR   WPI; 2004-081723/08.
DR   P-PSDB; ADJ11356.
XX
PT   New isolated S2 serine protease nucleic acids and polypeptides, useful
PT   for diagnosing and/or treating diseases with aberrant expression or
PT   activity the S2 serine protease, such as osteoarthritis, stress and
PT   apoptotic disorders.

```
XX
PS   Claim 2; SEQ ID NO 1; 28pp; English.
XX
CC   This invention relates to a novel isolated nucleic acid molecule
CC   identified as PRSS11-L that encodes an S2 serine protease. Specifically,
CC   it refers to members of the S2/HtrA serine protease family, such that it
CC   plays a role in cellular physiology and apoptosis. The present invention
CC   provides agonists, antagonists, antibodies and recombinant expression
CC   vectors useful in methods of treatment, or detection and diagnosis of
CC   diseases associated with the aberrant expression or activity of the S2
CC   serine protease, PRSS11-L. According to the invention described herein can
CC   be used via gene therapy routes to treat osteoarthritis, stress and
CC   apoptotic disorders. As such, they exhibit osteopathic, antiarthritic and
CC   tranquiliser activities. This polynucleotide sequence is the human PRSS11
CC   -L cDNA sequence of the invention.
XX
SQ   Sequence 3006 BP; 615 A; 918 C; 911 G; 562 T; 0 U; 0 Other;

Query Match             100.0%;  Score 1038;  DB 12;  Length 3006;
Best Local Similarity   100.0%;
Matches 1038;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 CAGGGACTCGAAGTTTGCAGTTCGTCCTCCACACTCAGTTCCACAGATGTGGTAGGAGGGGCAT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGGGACTCGAAGTTTGCAGTTCGTCCTCCACACTCAGTTCCACAGATGTGGTAGGAGGGGCAT 60

Qy         61 ATTCAGTCCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAAATCTGTCTGA 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTCAGTCCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAACGTAAGTAAATCTGTCTGA 120

Qy        121 GGCCACACAGCTAGAAAAGCAGCAGCCCAGCCCAGCCCAGAACCCCTGTGTGTTGTGCAGCC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCACACAGCTAGAAAAGCAGCAGCCCAGCCCAGCCCAGAACCCCTGTGTGTTGTGCAGCC 180

Qy        181 CCAGTTGCTCATTGCCGGGGCTCCGGAGCCCACGAGGCCATGTGCTTCCAGA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGTTGCTCATTGCCGGGGCTCCGGAGCCCACGAGGCCATGTGCTTCCAGA 240

Qy        241 TGGTGGGAACTGGAGAGAGGAGCCCGGCACAGGCCACCCGGAGGAACCCCGGAGGGCGTAGGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGGTGGGAACTGGAGAGAGGAGCCCGGCACAGGCCACCCGGAGGAACCCCGGAGGGCGTAGGC 300

Qy        301 CCCGTGCCACTGCATGCCCTCAGGCCCTGTGGTCCTGGCAGCCCACAGAGCCCCGGCCTGACG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCGTGCCACTGCATGCCCTCAGGCCCTGTGGTCCTGGCAGCCCACAGAGCCCCGGCCTGACG 360

Qy        361 GCAGCAGGAATCTGAGCCCCGGAGGGTCCAGGGAGTTCGTGAACCATTCTAGGAAGTCG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCAGCAGGAATCTGAGCCCCGGAGGGTCCAGGGAGTTCGTGAACCATTCTAGGAAGTCG 420

Qy        421 GGCTGGGGGTGTGGCCAAGTTAGACACAGAGATGTAGGGGCCCCTGTGGACTCAGAAAATTGGGCAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCTGGGGGTGTGGCCAAGTTAGACACAGAGATGTAGGGGCCCCTGTGGACTCAGAAAATTGGGCAG 480

Qy        481 CTCTTTTTGCCCAGAGGGGCCACCGTGTCCGGGCCCTGGGTAGCTCAGAAGGGTCACCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTCTTTTTGCCCAGAGGGGCCACCGTGTCCGGGCCCTGGGTAGCTCAGAAGGGTCACCT 540

Qy        541 GGGGGTCTTCCACTACACCCCCCGGCTGGACACTCTGCTGGGGCTCGGAGGGACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGGGTCTTCCACTACACCCCCCGGCTGGACACTCTGCTGGGGCTCGGAGGGACC 600

Qy        601 AGTGGAGCCCATGAGGAGGAGGGCCAGTTCTCTCCCGTAAGGGTATTGCTGTAGCACATGAG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGTGGAGCCCATGAGGAGGAGGGCCAGTTCTCTCCCGTAAGGGTATTGCTGTAGCACATGAG 660

Qy        661 GGAACAGAAGGCCCAGGGGGACTAACCCGGGGGCCTCACTCCCGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGAACAGAAGGCCCAGGGGGACTAACCCGGGGGCCTCACTCCCGTGTG 720

Qy        721 GCTCACGGCAAATACCTAACCTCTCTCTGAGCCCCTGGCCTAGCAGGGTTCCAGTG 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTCACGGCAAATACCTAACCTCTCTCTGAGCCCCTGGCCTAGCAGGGTTCCAGTG 780

Qy        781 AGGGGGGGTGAGGAAGCCCAGCACGTGGAAGCCTTTTTAACCATTCTCGGGGGTGAGCGAGC 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGGGGGGTGAGGAAGCCCAGCACGTGGAAGCCTTTTTAACCATTCTCGGGGGTGAGCGAGC 840

Qy        841 CCCTTCCCAAATGCCTGGTGTCACTGCACTGCTTGGTAGGGGTCCCCAACGGGCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCTTCCCAAATGCCTGGTGTCACTGCACTGCTTGGTAGGGGTCCCCAACGGGCTC 900

Qy        901 AGTGTGGGCTGAGGCTGGCTCTGAAACTGGGACACAGGGGGTCTCAGGAAGAGAGCCTTCCTCC 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTGTGGGCTGAGGCTGGCTCTGAAACTGGGACACAGGGGGTCTCAGGAAGAGAGCCTTCCTCC 960

Qy        961 TGCCCACTGGGCATAGGCCTCTCGGGAGCTGGGAGGGGTGGGCGCATGCAGCCATGCGTGAGCTCACTGATGCACCTGG 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGCCCACTGGGCATAGGCCTCTCGGGAGCTGGGAGGGGTGGGCGCATGCAGCCATGCGTGAGCTCACTGATGCACCTGG 1020

Qy       1021 CCCTTCCCGCCAGGCGCAG 1038
              |||||||||||||||||||
Db       1021 CCCTTCCCGCCAGGCGCAG 1038


RESULT 2
AAF21695/c
ID   AAF21695 standard; DNA; 1379 BP.
XX
AC   AAF21695;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 82.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; de.
OS   Homo sapiens.
XX
PN   WO20005173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US005881.
XX
PR   12-MAR-1999;  99US-0124270P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58792.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention, treatment
PT   and diagnosis of cancer, immune disorders, cardiovascular disorders and
PT   neurological diseases.
XX
PS   Claim 1; Page 543; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive; nootropic;
CC   neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
```

```
CC   antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
CC   antifungal; antiparasitic and cardiant activity. The polynucleotide and
CC   protein sequences are used in the diagnosis of cancer, particularly
CC   breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
CC   and agonists may also be used in the diagnosis, prevention and treatment
CC   of immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   cardiovascular disorders such as myocardial ischaemias; wound healing; and
CC   neurological diseases such as cerebral anoxia and epilepsy; and
CC   infectious diseases
XX
SQ   Sequence 1379 BP; 301 A; 354 C; 385 G; 330 T; 0 U; 9 Other;

Query Match             2.1%;  Score 22;  DB 3;  Length 1379;
Best Local Similarity   100.0%;  Pred. No. 5.5;
Matches  22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          120 AGGCCACACAGCTAGAAAGCAG 141
                |||||||||||||||||||||||
Db          790 AGGCCACACAGCTAGAAAGCAG 769

RESULT 3
ID   ADD47518 standard; DNA; 4558 BP.
XX
AC   ADD47518;
XX
DT   02-DEC-2004  (revised)
DT   29-JAN-2004  (first entry)
XX
DE   Human gene XM_045474, SEQ ID NO 13214.
XX
KW   Human; ds; gene; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
OS   Unidentified.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; XM_045474.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Example 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
```

```
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition. The polynucleotide or the
CC   polypeptides or their antibodies, for preparing a medicament for treating
CC   modulates its activity is useful for identifying the one or more
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human DNA (described in Table 3 of
CC   the specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4558 BP; 1030 A; 1062 C; 1303 G; 1163 T; 0 U; 0 Other;

Query Match             2.1%;  Score 22;  DB 10;  Length 4558;
Best Local Similarity   100.0%;  Pred. No. 5.2;
Matches  22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          772 GGTCCAGTGAGGGGGGGTGAGGA 793
                |||||||||||||||||||||||
Db         4145 GGTCCAGTGAGGGGGGGTGAGGA 4166

RESULT 4
ID   ADD47514 standard; DNA; 4558 BP.
XX
AC   ADD47514;
XX
DT   02-DEC-2004  (revised)
DT   29-JAN-2004  (first entry)
XX
DE   Human gene XM_045474, SEQ ID NO 13210.
XX
KW   Human; ds; gene; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
OS   Unidentified.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; XM_045474.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Example 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
```

```
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound
CC   modulates its activity of the compound that regulates the activity of
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic construction
CC   injury (CCI) and spred nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human DNA (described in Table 3 of
CC   the specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequence.
XX
SQ   Sequence 4558 BP; 1030 A; 1062 C; 1303 G; 1163 T; 0 U; 0 Other;

Query Match             2.1%;  Score 22;  DB 10;  Length 4558;
Best Local Similarity   100.0%;  Pred. No. 5.2;
Matches   22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   772  GGTCCAGTGAGGGGGGTGAGGA  793
          ||||||||||||||||||||||
Db  4145  GGTCCAGTGAGGGGGGTGAGGA  4166

RESULT 5
ID   ABD32665/c
ID   ABD32665 standard; DNA; 112453 BP.
XX
AC   ABD32665;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD13-070.
XX
KW   Human; d8; cancer-associated protein; gene; cyostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-0038838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW, Morris DW, Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukaemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 209; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 112453 BP; 28299 A; 24810 C; 27274 G; 32070 T; 0 U; 0 Other;

Query Match             2.1%;  Score 22;  DB 13;  Length 112453;
Best Local Similarity   100.0%;  Pred. No. 4.6;
Matches   22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    113  CTGTCTGAGGGCCACACAGCTAG  134
           |||||||||||||||||||||||
Db  33503  CTGTCTGAGGGCCACACAGCTAG  33482

RESULT 6
ID   ADJ11359
ID   ADJ11359 standard; DNA; 34 BP.
XX
AC   ADJ11359;
XX
DT   20-MAY-2004  (first entry)
XX
DE   PCR primer used to construct the human PRSS11-L expression vector SeqID5.
XX
KW   human; PRSS11-L; PCR; 88; S2 serine protease; S2/HtrA;
KW   cellular physiology; apoptosis; gene therapy; osteoarthritis; stress;
KW   apoptotic; osteopathic; antiarthritic; tranquiliser; primer.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US2004005659-A1.
XX
PD   08-JAN-2004.
XX
PF   03-JUL-2002; 2002US-00189099.
XX
```

PR   03-JUL-2002; 2002US-0189099.
XX
PA   (DARR/) DARROW A L.
PA   (QLJJ/) QI J.
PA   (CHEN/) CHEN C.
PA   (ANDR/) ANDRADE-GORDON P.
XX
PI   Darrow AL, Qi J, Chen C, Andrade-Gordon P;
XX
DR   WPI; 2004-081723/08.
XX
PT   New isolated S2 serine protease nucleic acids and polypeptides, useful
PT   for diagnosing and/or treating diseases with aberrant expression or
PT   activity the S2 serine protease, such as osteoarthritis, stress and
PT   apoptotic disorders.
XX
PS   Example 3; SEQ ID NO 5; 28pp; English.
XX
CC   This invention relates to a novel isolated nucleic acid molecule
CC   identified as PRSS11-L that encodes an S2 serine protease. Specifically,
CC   it refers to members of the S2/HtrA serine protease family, such that it
CC   plays a role in cellular physiology and apoptosis. The present invention
CC   provides agonists, antagonists, antibodies and recombinant expression
CC   vectors useful in methods of treatment, or detection and diagnosis of
CC   diseases associated with the aberrant expression or activity of the S2
CC   serine protease, PRSS11-L. Accordingly, compositions described herein can
CC   be used via gene therapy routes to treat osteoarthritis, stress and
CC   apoptotic disorders. As such, they exhibit osteopathic, antiarthritic and
CC   tranquiliser activities. This oligonucleotide sequence is a PCR primer
CC   used to construct the recombinant expression vector for human PRSS11-L
CC   cDNA of the invention.
XX
SQ   Sequence 34 BP; 4 A; 15 C; 8 G; 7 T; 0 U; 0 Other;

Qy   1011 ATGCACCTGGCCCTTCCCGCC 1031
               |||||||||||||||||||||
Db   14 ATGCACCTGGCCCTTCCCGCC 34

Query Match   2.0%; Score 21; DB 12; Length 34;
Best Local Similarity   100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
ADG99925/c
ID   ADG99925 standard; DNA; 244 BP.
XX
AC   ADG99925;
XX
DT   11-MAR-2004 (first entry)
XX
DE   Kidney disease-associated gene-related mouse DNA sequence SeqID1055.
XX
KW   kidney disease; proximal tubule; nephrotropic; chronic renal failure;
KW   diabetic renopathy; mouse; murine; ds.
XX
OS   Mus sp.
XX
PN   WO2003091427-A1.
XX
PD   06-NOV-2003.
XX
PF   23-APR-2003; 2003WO-JP005137.
XX
PR   24-APR-2002; 2002JP-0122253.
XX
PA   (KANS-) KANSAI TECHNOLOGY LICENSING ORG CO LTD.
XX
PI   Takenaka M, Imai E, Okubo K;
XX
DR   WPI; 2004-011769/01.
XX
PT   Kidney disease-associated genes expressed in proximal tubule, applicable

PT   in diagnosis of, in drug development and in devising appropriate strategy
PT   in treatment of e.g. chronic renal failure and diabetic renopathy.
XX
PS   Claim 1; SEQ ID NO 1055; 770pp; Japanese.
XX
CC   This invention relates to novel kidney disease-associated genes expressed
CC   in proximal tubules. The invention may be useful for the development of
CC   compounds with a nephrotropic activity. The genes are applicable in the
CC   diagnosis of, in drug development and in devising appropriate strategy in
CC   the treatment of (for example) chronic renal failure and diabetic
CC   renopathy. The present sequence is that of a mouse DNA sequence which is
CC   related to the invention.
XX
SQ   Sequence 244 BP; 55 A; 67 C; 65 G; 56 T; 0 U; 1 Other;

Qy   903 TGTGGGCGTGAGGCGTGGCTGG 923
               ||||||-|||||||||||||||
Db   102 TGTGGGCGTGAGGCGTGGCTG 82

Query Match   2.0%; Score 21; DB 12; Length 244;
Best Local Similarity   100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
ABL99890/c
ID   ABL99890 standard; cDNA; 1795 BP.
XX
AC   ABL99890;
XX
DT   03-OCT-2002 (first entry)
XX
DE   Human secretory polynucleotide (sptm) 145.
XX
KW   Human; ssi; gene; secretory protein; secretory polynucleotides; SPTM;
KW   SPTM-related disease; somatic gene therapy; germline gene therapy;
KW   severe combined immunodeficiency; intracellular parasite protection;
KW   fungal parasite; protozoan parasite; cell proliferative disorder; cancer;
KW   immune disorder; AIDS; neurological disorder; Parkinson's disease;
KW   motor neuron disorder; demyelinating disease; multiple sclerosis;
KW   meningitis; abscess; prion diseases; cerebral palsy;
KW   neuroskeletal disorder; peripheral nervous system disorder;
KW   dermatomyositis; polymyositis; myopathy; myasthenia gravis;
KW   mental disorder; Tourette's syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200220756-A2.
XX
PD   14-MAR-2002.
XX
PF   30-AUG-2001; 2001WO-US027297.
XX
PR   05-SEP-2000; 2000US-0229747P.
PR   05-SEP-2000; 2000US-0229748P.
PR   05-SEP-2000; 2000US-0229749P.
PR   05-SEP-2000; 2000US-0229750P.
PR   05-SEP-2000; 2000US-0229751P.
PR   05-SEP-2000; 2000US-0230016P.
PR   05-SEP-2000; 2000US-0230583P.
PR   05-SEP-2000; 2000US-0230505P.
PR   05-SEP-2000; 2000US-0230514P.
PR   05-SEP-2000; 2000US-0230515P.
PR   05-SEP-2000; 2000US-0230517P.
PR   05-SEP-2000; 2000US-0230518P.
PR   05-SEP-2000; 2000US-0230519P.
PR   06-SEP-2000; 2000US-0230595P.
PR   06-SEP-2000; 2000US-0230596P.
PR   06-SEP-2000; 2000US-0230597P.
PR   06-SEP-2000; 2000US-0230599P.
PR   06-SEP-2000; 2000US-0230600P.
PR   06-SEP-2000; 2000US-0230610P.
PR   06-SEP-2000; 2000US-0230864P.
PR   06-SEP-2000; 2000US-0308064P.
PR   06-SEP-2000; 2000US-0308065P.

```
PR   06-SEP-2000; 2000US-0230988P.
PR   06-SEP-2000; 2000US-0230989P.
PR   06-SEP-2000; 2000US-0230990P.
PR   07-SEP-2000; 2000US-0230896P.
PR   07-SEP-2000; 2000US-0230897P.
PR   07-SEP-2000; 2000US-0230951P.
PR   07-SEP-2000; 2000US-0231163P.
PR   07-SEP-2000; 2000US-0231832P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI   Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI   Moriyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI   Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
PI   Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
PS   Claim 1; Page 336-337; 585pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of human
CC   secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are
CC   useful for treating a disease or condition associated with the expression
CC   of functional SPTM. The SPTM DNA sequences are useful for somatic or
CC   germline gene therapy to correct a genetic deficiency (e.g. severe
CC   combined immunodeficiency). The SPTM DNA sequences are also useful in
CC   providing protection against intracellular parasites (e.g. fungal
CC   parasites and protozoan parasites). The SPTM DNA and protein sequences
CC   are also useful for diagnosing cell proliferative disorders, cancer,
CC   immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's
CC   disease), motor neuron disorders, demyelinating diseases (e.g. multiple
CC   sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,
CC   neuroskeletal disorders, peripheral nervous system disorders,
CC   dermatomyositis and polymyositis, myopathy, myasthenia gravis, and mental
CC   disorders (e.g. Tourette's syndrome). cDNA sequences ABL99746 - ABL99929
CC   represent human secretory polynucleotides of the invention
XX
SQ   Sequence 1795 BP; 350 A; 585 C; 502 G; 358 T; 0 U; 0 Other;

Qy          1006 CACTGATGCACCTGGCCCTTC 1026
                 |||||||||||||||||||||
Db           624 CACTGATGCACCTGGCCCTTC 604
```

```
RESULT 9
ID   ADQ87121/c
AC   ADQ87121;
XX
DT   07-OCT-2004 (first entry)
XX
DE   Human tumour-associated antigenic target (TAT) cDNA sequence #3997.
XX
KW   human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;
KW   cancer; cell proliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200406027O-A2.
XX
PD   22-JUL-2004.
XX
```

```
PF   15-OCT-2003; 2003WO-US029126.
XX
PR   18-OCT-2002; 2002US-0418988P.
XX
PA   (GETH ) GENENTECH INC.
PA   (WUTD/) WU T D.
PA   (ZHOU/) ZHOU Y.
XX
PI   Wu TD, Zhou Y;
XX
DR   WPI; 2004-534300/51.
XX
PS   Claim 1; SEQ ID NO 3997; 5504pp; English.
XX
CC   The present invention describes an isolated tumour-associated antigenic
CC   target (TAT) nucleic acid comprising: (a) any of the 4622 nucleotide
CC   sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of
CC   (a); (c) the complement of (a) or (b); (d) a sequence that has 80%
CC   sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-
CC   (c). Also described: (1) an expression vector comprising the above
CC   nucleic acid; (2) a host cell comprising the above expression vector; (3)
CC   a process for producing a polypeptide; (4) an isolated polypeptide
CC   comprising: (a) an amino acid sequence encoded by any of the above
CC   nucleotide sequences; (b) an amino acid sequence encoded by the full-
CC   length coding region of the above nucleotide sequences; or (c) a sequence
CC   having at least 80% identical to (a) or (b); (5) a chimeric polypeptide
CC   comprising the above polypeptide fused to a heterologous polypeptide; (6)
CC   an isolated antibody that binds to the above polypeptide; (7) a process
CC   for producing the antibody; (8) an isolated oligopeptide that binds to
CC   the above polypeptide; (9) a tumour-associated antigenic target (TAT)
CC   binding organic molecule that binds to the above polypeptide; (10) a
CC   composition of matter comprising the above (chimeric) polypeptide,
CC   antibody, oligopeptide or TAT binding organic molecule, in combination
CC   with a carrier; (11) an article of manufacture comprising a container and
CC   the composition of matter contained within the container; (12) methods of
CC   inhibiting the growth of a cell that expresses the above protein, where
CC   the growth of the cell is at least in part dependent upon a growth
CC   potentiating effect of the above protein; (13) a method of
CC   therapeutically treating a mammal having a cancerous tumour comprising
CC   cells that express the above protein; (14) a method of determining the
CC   presence of a protein in a sample suspected of containing the protein
CC   described above; (15) methods of diagnosing the presence of a tumour in a
CC   mammal; (16) a method for treating or preventing a cell proliferative
CC   disorder associated with increased expression or activity of the above
CC   protein; and (17) a method of binding an antibody, oligopeptide or
CC   organic molecule to a cell that expresses the protein described above.
CC   The TAT sequences have cytostatic activities, and can be used in gene
CC   therapy. The composition and methods are useful for diagnosing,
CC   preventing or treating cancer. The composition is also used for preparing
CC   a medicament for the therapeutic treatment or diagnostic detection of a
CC   cell proliferative disorder or cancer. The present sequence represents a
CC   human TAT cDNA sequence from the present invention.
XX
SQ   Sequence 3157 BP; 636 A; 1026 C; 812 G; 683 T; 0 U; 0 Other;

Qy          1006 CACTGATGCACCTGGCCCTTC 1026
                 |||||||||||||||||||||
Db           541 CACTGATGCACCTGGCCCTTC 521
```

```
RESULT 10
ID   ABZ36074/c
AC   ABZ36074;
XX
```

```
Qy    Best Local Match:  2.0%;   Score 21;  DB 6;   Length 1795;
      Matches   21;  Conservative   0;  Mismatches  16;  Indels  0;  Gaps  0;
Query Match           100.0%;  Pred. No. 16;
```

```
Db    ADQ87121 standard; cDNA; 3157 BP.
```

```
Qy    Best Local Similarity:  2.0%;  Score 21;  DB 12;  Length 3157;
      Matches   21;  Conservative   0;  Mismatches  16;  Indels  0;  Gaps  0;
Query Match           100.0%;  Pred. No. 16;
```

```
Db    ABZ36074 standard; cDNA; 3983 BP.
```

```
DT   10-FEB-2003 (first entry)
XX
DE   Human secretory polynucleotide SPTM SEQ ID NO 238.
XX
KW   Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
KW   asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
KW   Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disorder;
KW   multiple sclerosis; Parkinson's disease; cell proliferative disorder;
KW   anti-inflammatory; immunosuppressive; neuroprotective; nootropic;
KW   neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;
KW   antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;
KW   secretory polynucleotide; secretory protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20028376-A2.
XX
PD   24-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-US009921.
XX
PR   29-MAR-2001; 2001US-0280067P.
PR   29-MAR-2001; 2001US-0280068P.
PR   16-MAY-2001; 2001US-0291280P.
PR   17-MAY-2001; 2001US-0291829P.
PR   17-MAY-2001; 2001US-0291849P.
PR   19-JUN-2001; 2001US-0299428P.
PR   20-JUN-2001; 2001US-0299776P.
PR   20-JUN-2001; 2001US-0300001P.
XX
XX
DR   WPI; 2003-075543/07.
DR   P-PSDB; ABP75630.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI   Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
PI   Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI   Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI   Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
PT   New human secretory proteins and polynucleotides, useful for diagnosing,
PT   treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),
PT   neurological disorders (e.g. Alzheimer's), or cell proliferations or
PT   cancers.
XX
PS   Claim 1; SEQ ID NO 238; 458pp + Sequence Listing; English.
XX
CC   The invention relates to a secretory polynucleotide (designated sptm)
CC   comprising any of 567 polynucleotide sequences (AB235837-AB236403), a
CC   naturally occurring polynucleotide sequence at least 90 % identical to
CC   the polynucleotide sequence, a polynucleotide complementary to them or an
CC   RNA equivalent of them. The polypeptide or polynucleotide are useful for
CC   treating, preventing or diagnosing a disease or condition associated with
CC   the expression of functional SPTM. These are particularly useful for
CC   diagnosing, treating or preventing autoimmune/inflammatory disorders
CC   (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
CC   disease), neurological disorders (e.g. epilepsy, Huntington's disease,
CC   dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
CC   multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
CC   schizophrenia or amnesia) or cell proliferative disorders (e.g.
CC   psoriasis, lymphoma, myeloma, sarcoma or cancers including adenocarcinoma,
CC   leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
CC   breast, cervix or prostate). Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3983 BP; 819 A; 1257 C; 980 G; 924 T; 0 U; 3 Other;

Query Match             2.0%;
Best Local Similarity   100.0%;  Score 21;  DB 8;  Length 3983;
Matches   21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
                                 Pred. No. 16;
```

```
Qy   1006 CACTGATGCACCTGGCCCTTC 1026
           ||||||||||||||||||||||
Db   624  CACTGATGCACCTGGCCCTTC 604

RESULT 11
ADM87103/c
ID   ADM87103 standard; cDNA; 4030 BP.
XX
AC   ADM87103;
XX
DT   03-JUN-2004 (first entry)
XX
DE   Human protein encoding cDNA SEQ ID NO:196.
XX
KW   respiratory; cytostatic; antiarthritic; antiinflammatory;
KW   gastrointestinal; antibacterial; immunosuppressive; antidiabetic;
KW   antirheumatic; gene therapy; molecular weight marker; chromosome marker;
KW   chromosome tag; genetic fingerprinting; nutritional supplement; cancer;
KW   inflammatory condition; arthritis; inflammatory bowel disease;
KW   Crohn's disease; sepsis; rheumatoid arthritis; diabetes mellitus type 1;
KW   graft versus host disease; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200409834-A2.
XX
PD   29-JAN-2004.
XX
PF   19-JUL-2002; 2002WO-US022858.
XX
PR   21-JUL-2001; 2001US-0306971P.
PR   28-MAR-2002; 2002US-00112944.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Tang YT, Yang Y, Weng G, Zhang J, Ren F, Xue A, Wang J;
PI   Wehrman T, Ghosh MJ, Wang D, Zhao QA, Wang Z;
XX
DR   WPI; 2004-143291/14.
DR   P-PSDB; ADM87347.
XX
PT   New isolated polynucleotides and polypeptides, useful for treating, e.g.
PT   cancer, lung or liver fibrosis, arthritis, inflammatory bowel disease,
PT   Crohn's disease, rheumatoid arthritis, diabetes mellitus type 1 or graft
PT   versus host disease.
XX
PS   Claim 1; SEQ ID NO 196; 591pp; English.
XX
CC   The present invention describes an isolated polynucleotide (I): (a)
CC   comprising a nucleotide sequence selected from SEQ ID NO:1-244; or (b)
CC   which encodes a polypeptide with biological activity, where the
CC   polynucleotide hybridises to (I) under stringent hybridisation conditions
CC   or has greater than 99% sequence identity with (I). (I) has respiratory,
CC   cytostatic, antiarthritic, antiinflammatory, gastrointestinal,
CC   antibacterial, immunosuppressive, antidiabetic and antirheumatic
CC   activities, and can be used in gene therapy. (I) can be used for
CC   generating polynucleotides encoding chimeric or fusion proteins and
CC   heterologous protein sequences. The polynucleotides can be used to
CC   express recombinant protein for analysis, characterisation or therapeutic
CC   use; as markers for tissues in which the corresponding protein is
CC   preferentially expressed; as molecular weight markers on gels; as
CC   chromosome markers or tags to identify chromosomes or to map related gene
CC   positions; to compare with endogenous DNA sequences in patients to
CC   identify potential genetic disorders; as probes to hybridise and discover
CC   genes, related DNA sequences; as a source of information to derive PCR
CC   primers for genetic fingerprinting; as a probe to subtract-out known
CC   sequences in the process of discovering other novel polynucleotides; for
CC   selecting and making oligomers for attachment to a gene chip or other
CC   support, including for examination of expression patterns; to raise anti-
CC   protein antibodies using DNA immunisation techniques; and as an antigen
CC   to raise anti-DNA antibodies or elicit another immune response. The
CC   polynucleotides and polypeptides can also be used as nutritional sources
```

CC   or supplements, e.g. as a protein or amino acid supplement, as a carbon
CC   source, as a nitrogen source or as a source of carbohydrates. The
CC   polynucleotides and polypeptides can also be used treat cancer. The
CC   compositions are useful for promoting better or faster closure of non-
CC   healing wounds, for the generation and regeneration of tissues, for gut
CC   protection or regeneration of lung or liver fibrosis,
CC   reperfusion injury in various tissues, and conditions resulting from
CC   systemic cytokine damage. The compositions can also be used to treat
CC   inflammatory conditions (e.g. arthritis, inflammatory bowel disease or
CC   Crohn's disease), sepsis, rheumatoid arthritis, diabetes mellitus type 1
CC   or graft versus host disease. The present sequence represents a novel
CC   human polynucleotide sequence from the present invention. N.B. The
CC   sequences for this patent obtained from the USPTO web site from an
CC   equivalent US patent US2004048249A1.
XX
SQ   Sequence 4030 BP; 828 A; 1269 C; 1056 G; 876 T; 0 U; 1 Other;

Qy   1006 CACTGATGCAACCTGGCCCCTTC 1026
          |||||||||||||||||||||||
Db   1508 CACTGATGCAACCTGGCCCCTTC 1488

RESULT 12

ID   ADQ80254 standard; cDNA; 127145 BP.
AC   ADQ80254;
XX
DT   21-OCT-2004 (first entry)
XX
DE   Hermansky-Pudlak syndrome associated cDNA.
XX
KW   ss; gene; cytostatic; epidermal growth factor receptor modulator;
KW   identification; therapeutic response; cancer; EGFR; biomarker.
XX
OS   Homo sapiens.
XX
PN   WO2004063709-A2.
XX
PD   29-JUL-2004.
XX
PF   08-JAN-2004; 2004WO-US000368.
XX
PR   08-JAN-2003; 2003US-0438735P.
XX
PA   (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI   Amler LC, Januario T;
XX
DR   WPI; 2004-544114/52.
DR   P-PSDB; ADQ80373.
XX
PS   Disclosure; SEQ ID NO 26; 520pp; English.

PT   The invention relates to a method of identifying a mammal that will
PT   respond therapeutically to a method of treating cancer by administering
PT   an epidermal growth factor receptor (EGFR) modulator by comparing the
PT   level of a biomarker in a mammal before and after exposure to an EGFR
PT   modulator.

CC   Identifying a mammal that will respond therapeutically to a method of
CC   treating cancer comprises comparing the level of a biomarker in a mammal
CC   before and after exposure to an epidermal growth factor receptor (EGFR)
CC   modulator.

Query Match   2.0%; Score 21; DB 12; Length 4030;
Best Local Similarity 100.0%;   Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC   biomarkers are useful for identifying a mammal that will respond
CC   therapeutically to a method of treating cancer by administering an
CC   epidermal growth factor receptor (EGFR) modulator. This sequence
CC   corresponds to one of the biomarkers whose levels of gene expression is
CC   measured in the method of the invention.
XX
SQ   Sequence 127145 BP; 31066 A; 31499 C; 32828 G; 31752 T; 0 U; 0 Other;

Qy   1006 CACTGATGCAACCTGGCCCCTTC 1026
          |||||||||||||||||||||||
Db   58197 CACTGATGCAACCTGGCCCCTTC 58217

Query Match   2.0%; Score 21; DB 13; Length 127145;
Best Local Similarity 100.0%;   Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13

ID   ACF62741/c
ID   ACF62741/c standard; DNA; 189013 BP.
AC   ACF62741;
XX
DT   08-OCT-2003 (first entry)
XX
DE   Cancer based on CYP3A5 related polynucleotide SEQ ID NO:669.
XX
KW   Cancer; CYP3A5; irinotecan; pharmaceutical; malignant glioma;
KW   cytochrome p450; subfamily IIIA; nifedipine oxidase; polypeptide 5;
KW   cytostatic; gene; ds.
XX
OS   Unidentified.
XX
PN   WO2003013534-A2.
XX
PD   20-FEB-2003.
XX
PF   23-JUL-2002; 2002WO-EP008219.
XX
PR   23-JUL-2001; 2001EP-00117608.
PR   24-MAY-2002; 2002EP-00011710.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Heinrich G, Kerb R;
XX
DR   WPI; 2003-268144/26.
XX
PS   Disclosure; SEQ ID NO 669; 86pp; English.

PT   New use of irinotecan for preparation of compositions for treating cancer
PT   in subject having genome with variant allele comprising cytochrome p450,
PT   subfamily IIIA, polypeptide 5 polynucleotide, termed CYP3A5.

CC   The present invention describes the use of irinotecan (I) or its
CC   derivative for the preparation of a pharmaceutical composition for
CC   treating colorectal, cervical, gastric, lung, ovarian or pancreatic
CC   cancer, or malignant glioma in a subject having a genome with a variant
CC   allele which comprises a cytochrome p450, subfamily IIIA (nifedipine
CC   oxidase), polypeptide 5 (CYP3A5) polynucleotide (II). (I) and (II) have
CC   cytostatic activity. The therapeutic applications of (II) is improved,
CC   since it is possible to individually treat a subject with an appropriate
CC   dosage and/or an appropriate derivative of (I). Therefore, undesirable,
CC   harmful or toxic effects are efficiently avoided. Unnecessary and
CC   potentially harmful treatment of those subjects who do not respond to the
CC   treatment with substances (nonresponders), as well as the development of
CC   drug resistances due to suboptimal drug dosing can be avoided. ACF62200
CC   to ACF62751 and ABM34912 to ABM35013 represent sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 189013 BP; 53603 A; 38975 C; 39348 G; 57087 T; 0 U; 0 Other;

Query Match   2.0%; Score 21; DB 8; Length 189013;

```
RESULT 14
ADB20856/c
ID   ADB20856 standard; DNA; 189013 BP.
XX
AC   ADB20856;
XX
Db   4162 TGTCTGAGGCCACACAGCTAG 4142
     ||||||||||||||||||||||
Qy    114 TGTCTGAGGCCACACAGCTAG 134
Matches 21; Conservative  0; Mismatches  0; Indels  0; Gaps  0;
Best Local Similarity  100.0%; Pred. No. 13;
```

```
DT   20-NOV-2003  (first entry)
XX
AC   ADB20856;
XX
DE   MRP1 based cancer related nucleic acid SEQ ID NO:669.
XX
KW   irinotecan; colorectal cancer; cervical cancer; gastric cancer;
KW   lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;
KW   variant allele; multidrug resistance protein 1; MRP1; cytostatic; gene;
KW   ds.
XX
OS   Unidentified.
XX
PN   WO200301353-A2.
XX
PD   20-FEB-2003.
XX
PF   23-JUL-2002; 2002WO-EP008200.
XX
PR   23-JUL-2001; 2001EP-00117608.
PR   24-MAY-2002; 2002EP-0001710.
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI   Heinrich G,  Kerb R;
XX
DR   WPI; 2003-354397/33.
XX
PT   Use of irinotecan or its derivative for preparation of a pharmaceutical
PT   composition for treating cancer in a subject having a genome with a
PT   variant allele comprising a multidrug resistance protein 1
PT   polynucleotide.
XX
PS   Disclosure; SEQ ID NO 669; 100pp; English.
XX
CC   The present invention describes a method for the use of irinotecan (I) or
CC   its derivative for the preparation of a pharmaceutical composition for
CC   treating colorectal, cervical, gastric, lung, ovarian or pancreatic
CC   cancer, or malignant glioma in a subject having a genome with a variant
CC   allele which comprises a multidrug resistance protein 1 (MRP1)
CC   polynucleotide (III). (I) has cytostatic activity. (II) or its derivative
CC   can be used for the preparation of a pharmaceutical composition for
CC   treating colorectal, cervical, gastric, lung, ovarian or pancreatic
CC   cancer, or malignant glioma in a subject, where the subject is a human
CC   (preferably African or Asian) or a mouse. The present sequence represents
CC   a sequence which is used in the exemplification of the present invention.
XX
SQ   Sequence 189013 BP; 53603 A; 38975 C; 39348 G; 57087 T; 0 U; 0 Other;

Qy    114 TGTCTGAGGCCACACAGCTAG 134
          ||||||||||||||||||||||
Db   4162 TGTCTGAGGCCACACAGCTAG 4142
Best Local Similarity  100.0%; Pred. No. 13;
Matches 21; Conservative  0; Mismatches  0; Indels  0; Gaps  0;
Query Match     2.0%;  Score 21;  DB 8;  Length 189013;
```

```
RESULT 15
ADB87945/c
ID   ADB87945 standard; DNA; 189013 BP.
XX
AC   ADB87945;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human UGT1A1 gene sequence SEQ ID NO:669.
XX
KW   irinotecan; cancer; UGT1A1; cytostatic; topoisomerase I inhibitor;
KW   colorectal cancer; cervical cancer; gastric cancer; lung cancer;
KW   ovarian cancer; pancreatic cancer; malignant glioma;
KW   uridine diphosphate glycosyltransferase1 member A1; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200301353-A2.
XX
PD   20-FEB-2003.
XX
DR   WPI; 2003-289896/28.
XX
PI   Heinrich G,  Kerb R;
XX
PA   (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PT   Use of irinotecan to treat cancer patient by determining if patient has
PT   variant alleles of UGT1A1 gene, administering increased/decreased amounts
PT   of irinotecan based on increased/decreased levels of UGT1A1 gene product.
XX
PR   23-JUL-2001; 2001EP-00117608.
PR   24-MAY-2002; 2002EP-00011710.
XX
PF   23-JUL-2002; 2002WO-EP008217.
XX
PS   Disclosure; SEQ ID NO 669; 107pp; English.
XX
CC   The invention relates to the novel use of irinotecan to treat a patient
CC   suffering from cancer. This involves determining if the patient has one
CC   or more variant alleles of the UGT1A1 gene, and if the patient has one or
CC   more of such variant alleles, irinotecan is administered in an increased
CC   or decreased amount in comparison to the amount that is administered
CC   without regard to the patient's alleles in the UGT1A1 gene. The invention
CC   has cytostatic activity. A composition of the invention acts as a
CC   topoisomerase I inhibitor. The method is useful for treating a patient,
CC   from cancer such as colorectal, cervical, gastric cancer, lung, ovarian,
CC   an animal e.g. mouse or a human, preferably African or Asian, suffering
CC   pancreatic cancer or malignant glioma. The present sequence is udes in
CC   the exemplification of the invention.
XX
SQ   Sequence 189013 BP; 53603 A; 38975 C; 39348 G; 57087 T; 0 U; 0 Other;

Qy    114 TGTCTGAGGCCACACAGCTAG 134
          ||||||||||||||||||||||
Db   4162 TGTCTGAGGCCACACAGCTAG 4142
Best Local Similarity  100.0%; Pred. No. 13;
Matches 21; Conservative  0; Mismatches  0; Indels  0; Gaps  0;
Query Match     2.0%;  Score 21;  DB 10;  Length 189013;

Search completed: June 29, 2005, 18:37:34
Job time : 686 secs
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd. GenCore version 5.1.6

Run on: June 29, 2005, 18:26:10 ; Search time 759 Seconds
(without alignments)
8548.139 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 cagggactcgaagttgcag..........ggccctccgccagcgcag 1038

Scoring table: Gapop 60.0 , Gapext 60.0
Searched: 6067389 seqs, 312525B755 residues

Word-size : 12

OLIGO_NUC.

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Total number of hits satisfying chosen parameters: 481239

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

		Description
1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*	
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*	
3:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*	
4:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*	
5:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*	
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*	
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*	
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*	
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*	
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*	
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*	
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*	
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*	
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*	
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*	
16:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*	
17:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*	
18:	/cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*	
19:	/cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*	
20:	/cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*	
21:	/cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*	
22:	/cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*	
23:	/cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*	
24:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*	
25:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*	
26:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*	

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1038	100.0	3006	17	US-10-189-099A-1	Sequence 1, Appli
2	1038	100.0	3006	21	US-10-617-443B-1	Sequence 1, Appli
3	22	2.1	1379	9	US-09-925-298-82	Sequence 82, Appli
4	22	2.1	1379	14	US-10-102-806-82	Sequence 82, Appl
5	22	2.1	113486	20	US-10-417-375-18	Sequence 18, Appl
6	21	2.0	34	17	US-10-189-0099A-5	Sequence 5, Appli
7	21	2.0	34	21	US-10-617-443B-5	Sequence 5, Appli

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

c	8	21	2.0	4030	18	US-10-112-944-196	Sequence 196, App
c	9	21	2.0	189013	21	US-10-484-577-669	Sequence 669, App
c	10	20	1.9	20	21	US-10-189-099A-7	Sequence 7, Appli
c	11	20	1.9	20	21	US-10-617-443B-7	Sequence 7, Appli
c	12	20	1.9	504	13	US-09-833-263-1058	Sequence 1058, Ap
c	13	20	1.9	504	13	US-09-873-367C-695	Sequence 695, App
c	14	20	1.9	1058	13	US-10-025-380-1058	Sequence 1058, Ap
c	15	20	1.9	15720	10	US-09-922-217-1058	Sequence 1058, Ap
c	16	20	1.9	15720	9	US-10-027-632-195771	Sequence 195771,
c	17	20	1.9	15720	17	US-10-393-590-14	Sequence 14, Appl
c	18	20	1.9	15720	16	US-10-393-567-14	Sequence 14, Appl
c	19	20	1.9	15720	16	US-10-394-087-13	Sequence 13, Appl
c	20	20	1.9	15720	16	US-10-394-087-14	Sequence 14, Appl
c	21	20	1.9	15720	19	US-10-734-564-54	Sequence 54, Appl
c	22	20	1.9	15720	21	US-10-887-553A-340	Sequence 340, App
c	23	20	1.9	15720	16	US-10-800-350-391	Sequence 391, App
c	24	20	1.9	15720	21	US-10-887-553A-340	Sequence 340, App
c	25	20	1.9	26000	21	US-10-765-790-74	Sequence 74, Appl
c	26	20	1.9	30542	22	US-10-322-696-139	Sequence 139, App
c	27	20	1.9	30542	22	US-10-737-082-74	Sequence 74, Appl
c	28	20	1.9	79860	13	US-10-087-192-412	Sequence 412, App
c	29	20	1.9	94529	17	US-10-034-650-52	Sequence 52, Appl
c	30	20	1.9	101365	20	US-10-719-993-6830	Sequence 6830, Ap
c	31	20	1.9	101365	20	US-10-719-993-6785	Sequence 6785, Ap
c	32	20	1.9	309996	9	US-1-719-993-6785	Sequence 6785, Ap
c	33	19	1.9	254	9	US-09-923-876-502	Sequence 502, App
c	34	19	1.8	254	10	US-09-923-876-502	Sequence 502, App
c	35	19	1.8	290	14	US-10-060-036-1290	Sequence 1290, Ap
c	36	19	1.8	336	14	US-10-060-036-1210	Sequence 1210, Ap
c	37	19	1.8	367	14	US-10-060-036-1044	Sequence 1044, Ap
c	38	19	1.8	410	14	US-10-060-036-740	Sequence 740, App
c	39	19	1.8	410	14	US-10-060-036-4528	Sequence 4528, A
c	40	19	1.8	424	20	US-10-357-930-53528	Sequence 53528, A
c	41	19	1.8	438	14	US-10-060-036-925	Sequence 925, App
c	42	19	1.8	440	14	US-10-060-036-836	Sequence 836, App
c	43	19	1.8	445	10	US-09-814-353-4536	Sequence 4536, Ap
c	44	19	1.8	445	10	US-09-814-353-10837	Sequence 10837, A
c	45	19	1.8	449	14	US-10-060-036-622	Sequence 622, App

ALIGNMENTS

RESULT 1
US-10-189-099A-1

Sequence 1, Application US/10189099A
Publication No. US2004005659A1

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew L
APPLICANT: Qi, Jian-shen
APPLICANT: Chen, Cailin
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
FILE REFERENCE: ORT-1644
CURRENT APPLICATION NUMBER: US/10/189,099A
CURRENT FILING DATE: 2002-12-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1
LENGTH: 3006
TYPE: cDNA
ORGANISM: Homo sapiens

US-10-189-099A-1

Query Match 100.0%;
Best Local Similarity 100.0%;
Matches 1038; Conservative 0;

Score 1038; DB 17; Length 3006;
Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CAGGGACTCGAAGTTGCAGTCCTCCACACTCAGTTCCCACACTCAGATGTGGTTAGGAGGCAT 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAGGGACTCGAAGTTGCAGTCCTCCACACTCAGTTCCCACACTCAGATGTGGTTAGGAGGCAT 60

Qy     61 ATTCAGTCCCATTTTTCAGATGAGGAGTTGAGGCCCAGAGAGAACGTAAGTAATCTGTCTGA 120
```

Query Match   100.0%;   Score 1038;   DB 21;   Length 3006;
Best Local Similarity   100.0%;   Pred. No. 0;   Indels 0;   Gaps 0;
Matches 1038;   Conservative   0;   Mismatches   0;

(DNA sequence alignment — Query (Qy) vs Database (Db), bases 1–1038, 100% identity)

```
Db   781  AGGGGGGGTGAGGAAGCCCAGCACGTGGAAGCCTTTTTAACCATTCTCGGGGGTGAGCGGAGC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   841  CCCTTCCCAAATGCCTGGTGTCACTCGCACTGCTCGTGTGTGGTAGGGGGTCCCAAACGGGCTC  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  CCCTTCCAAATGCCTGGTGTCACTCGCACTGCTCGTGTGTGGTAGGGGGTCCCAAACGGGCTC  900
Qy   901  AGTGTGGGGCTGAGGCTGGCTCTTGACTGGGACACAGGGGTCTCAGGGAGAGCCTTCCTCCTCC  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  AGTGTGGGGCTGAGGCTGGCTCTTGACTGGGACACAGGGGTCTCAGGGAGAGCCTTCCTCCTCC  960
Qy   961  TGCCCACTGGGCATAGGCCCTCTGGGAGGACTGGGCCAGCAGCATGGGGTTCACTGATGCACCTGG  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  TGCCCACTGGGCATAGGCCCTCTGGGAGGAGCTGGCCAGCAGCATCGGTGAATCTCACTGATGCACCTGG  1020
Qy  1021  CCCTTCCCGCCAGCGCAG  1038
          ||||||||||||||||||
Db  1021  CCCTTCCCGCCAGCGCAG  1038
```

RESULT 3
US-09-925-298-82/c
; Sequence 82, Application US/09925298
; Publication No. US20020039764A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103
; CURRENT APPLICATION NUMBER: US/09/925,298
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 82
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1365)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1378)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-298-82

  Query Match   2.1%; Score 22; DB 9; Length 1379;
  Best Local Similarity   100.0%;
  Matches   22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy   120  AGGCCACACAGCTAGAAAGCAG  141
          ||||||||||||||||||||||
Db   790  AGGCCACACAGCTAGAAAGCAG  769
```

RESULT 4
US-10-102-806-82/c
; Sequence 82, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 82
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1365)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1378)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-102-806-82

  Query Match   2.1%; Score 22; DB 14; Length 1379;
  Best Local Similarity   100.0%;
  Matches   22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy   120  AGGCCACACAGCTAGAAAGCAG  141
          ||||||||||||||||||||||
Db   790  AGGCCACACAGCTAGAAAGCAG  769
```

RESULT 5
US-10-417-375-18/c
; Sequence 18, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 113486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-18

  Query Match   2.1%; Score 22; DB 20; Length 113486;
  Best Local Similarity   100.0%;
  Matches   22; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    113  CTGTCTGAGGCCACACAGCAGCTAG  134
           |||||||||||||||||||||||||
Db  34495  CTGTCTGAGGCCACACAGCAGCTAG  34474
```

RESULT 6
US-10-189-099A-5
; Sequence 5, Application US/10189099A
; Publication No. US20040005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8

```
;  SOFTWARE: Patentin Ver. 2.1
;  SEQ ID NO 5
;  LENGTH: 34
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: primer
US-10-189-099A-5

Qy          1011 ATGCACCTGGCGCCCTTCCCGCC 1031
                 ||||||||||||||||||||||
Db            14 ATGCACCTGGCGCCCTTCCCGCC 34

Query Match          2.0%;  Score 21;  DB 17;  Length 34;
Best Local Similarity   100.0%;  Pred. No. 2.7;
Matches  21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;


RESULT 7
US-10-617-443B-5
Sequence 5, Application US/10617443B
Publication No. US20050019777A1
GENERAL INFORMATION:
;  APPLICANT: Darrow, Andrew L
;  APPLICANT: Qi, Jian-shen
;  APPLICANT: Chen, Cailin
;  APPLICANT: Andrade-Gordon, Patricia
;  TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
;  FILE REFERENCE: ORT-1644
;  CURRENT APPLICATION NUMBER: US/10/617,443B
;  CURRENT FILING DATE: 2003-07-02
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: Patentin Ver. 2.1
;  SEQ ID NO 5
;  LENGTH: 34
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: primer
US-10-617-443B-5

Qy          1011 ATGCACCTGGCGCCCTTCCCGCC 1031
                 ||||||||||||||||||||||
Db            14 ATGCACCTGGCGCCCTTCCCGCC 34

Query Match          2.0%;  Score 21;  DB 21;  Length 34;
Best Local Similarity   100.0%;  Pred. No. 2.7;
Matches  21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;


RESULT 8
US-10-112-944-196/c
Sequence 196, Application US/10112944
Publication No. US20040048249A1
GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Yang, Yonghong
;  APPLICANT: Weng, Gezhi
;  APPLICANT: Zhang, Jie
;  APPLICANT: Ren, Feiyan
;  APPLICANT: Xue, Aidong J.
;  APPLICANT: Wang, Jian-Rui
;  APPLICANT: Wehrman, Tom
;  APPLICANT: Ghosh, Malabika
;  APPLICANT: Wang, Dunrui
;  APPLICANT: Zhao, Qing A.
;  APPLICANT: Wang, Zhiwei
;  TITLE OF INVENTION: Secreted Polypeptides
;  TITLE OF INVENTION: No. US20040048249A1e1 Nucleic Acids and
;  FILE REFERENCE: 805A
;  CURRENT APPLICATION NUMBER: US/10/112,944
;  CURRENT FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 09/488,725
```

```
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: US 09/491,404
;  PRIOR FILING DATE: 2000-01-25
;  PRIOR APPLICATION NUMBER: US 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  PRIOR APPLICATION NUMBER: US 09/515,126
;  PRIOR FILING DATE: 2000-02-28
;  PRIOR APPLICATION NUMBER: US 09/519,705
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: US 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: US 09/552,929
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/577,408
;  PRIOR FILING DATE: 2000-05-18
;  NUMBER OF SEQ ID NOS: 924
;  SOFTWARE: pt_FL_genes Version 5.0
;  SEQ ID NO 196
;  LENGTH: 4030
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (985)..(3036)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(4030)
;  OTHER INFORMATION: n = a,t,c or g
US-10-112-944-196

Qy          1006 CACTGATGCACCTGGCGCCCTTC 1026
                 ||||||||||||||||||||||
Db          1508 CACTGATGCACCTGGCGCCCTTC 1488

Query Match          2.0%;  Score 21;  DB 18;  Length 4030;
Best Local Similarity   100.0%;  Pred. No. 1.7;
Matches  21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;


RESULT 9
US-10-484-577-669/c
Sequence 669, Application US/10484577
Publication No. US20050032724A1
GENERAL INFORMATION:
;  APPLICANT: EPIDAUROS Biotechnologie Aktiengesellschaft
;  TITLE OF INVENTION: Means and methods for improved treatment of cancer based on UGT1A1
;  FILE REFERENCE: F2285PCT-1
;  CURRENT APPLICATION NUMBER: US/10/484,577
;  CURRENT FILING DATE: 2004-01-22
;  PRIOR APPLICATION NUMBER: PCT/EP 02/08220
;  PRIOR FILING DATE: 2002-07-23
;  PRIOR APPLICATION NUMBER: EP 01 11 7608.8
;  PRIOR FILING DATE: 2001-07-23
;  PRIOR APPLICATION NUMBER: EP 02011710.7
;  PRIOR FILING DATE: 2002-05-24
;  NUMBER OF SEQ ID NOS: 683
;  SOFTWARE: Patentin version 3.1
;  SEQ ID NO 669
;  LENGTH: 189013
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-484-577-669

Qy           114 TGTCTGAGGCCACACAGCTAG 134
                 |||||||||||||||||||||
Db          4162 TGTCTGAGGCCACACAGCTAG 4142

Query Match          2.0%;  Score 21;  DB 21;  Length 189013;
Best Local Similarity   100.0%;  Pred. No. 1.2;
Matches  21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;


RESULT 10
```

US-10-189-0099A-7
; Sequence 7, Application US/10189099A
; Publication No. US20040005659A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/189,099A
; CURRENT FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 7
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Internal Sequence
; FEATURE:
; OTHER INFORMATION: Description of Internal Sequence: primer
US-10-189-0099A-7

Query Match   1.9%;   Score 20; DB 17; Length 20;
Best Local Similarity   100.0%;   Pred. No. 9.6;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   413 GCAAGTCGGCGCTGGGGTGTG 432
          ||||||||||||||||||||
Db   1 GCAAGTCGGCGCTGGGGTGTG 20

RESULT 11
US-10-617-443B-7
; Sequence 7, Application US/10617443B
; Publication No. US20050019777A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew L
; APPLICANT: Qi, Jian-shen
; APPLICANT: Chen, Cailin
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Human PRSS-11 like S2 serine protease and uses thereof
; FILE REFERENCE: ORT-1644
; CURRENT APPLICATION NUMBER: US/10/617,443B
; CURRENT FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 7
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Internal primer
US-10-617-443B-7

Query Match   1.9%;   Score 20; DB 21; Length 20;
Best Local Similarity   100.0%;   Pred. No. 9.6;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   413 GCAAGTCGGCGCTGGGGTGTG 432
          ||||||||||||||||||||
Db   1 GCAAGTCGGCGCTGGGGTGTG 20

RESULT 12
US-10-027-632-195771/c
; Sequence 195771, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SEQ ID NO 195771
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195771

Query Match   1.9%;   Score 20; DB 17; Length 504;
Best Local Similarity   100.0%;   Pred. No. 7.1;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   67 TCCCATTTTTCAGATGAGGA 86
          ||||||||||||||||||||
Db   464 TCCCATTTTTCAGATGAGGA 445

RESULT 13
US-10-027-632-195771/c
; Sequence 195771, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 195771
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195771

Query Match   1.9%;   Score 20; DB 13; Length 504;
Best Local Similarity   100.0%;   Pred. No. 7.1;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   67 TCCCATTTTTCAGATGAGGA 86
          ||||||||||||||||||||
Db   464 TCCCATTTTTCAGATGAGGA 445

```
RESULT 14
US-09-922-217-1058/c
; Sequence 1058, Application US/09922217
; Patent No. US2002007641A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Secrist, Heather
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Meagher, Madeleine Joy
;  APPLICANT: Stolk, John A.
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT: Smith, Carole Lynn
;  APPLICANT: King, Gordon E.
;  APPLICANT: Wang, Aijun
;  APPLICANT: Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
;  LENGTH: 15720
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-922-217-1058

Qy           316 GCCTCAGGCCCTGTGGTCCTG 335
                 ||||||||||||||||||||
Db         13006 GCCTCAGGCCCTGTGGTCCTG 12987

Query Match     1.9%;  Score 20;  DB 9;  Length 15720;
Best Local Similarity   100.0%;  Pred. No. 5.1;
Matches   20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 15
US-09-833-263-1058/c
; Sequence 1058, Application US/09833263
; Patent No. US2002110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT: Clapper, Jonathan D.
;  APPLICANT: Stolk, John A.
;  APPLICANT: Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE: 2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1058
;  LENGTH: 15720
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-833-263-1058

Qy           316 GCCTCAGGCCCTGTGGTCCTG 335
                 ||||||||||||||||||||
Db         13006 GCCTCAGGCCCTGTGGTCCTG 12987

Query Match     1.9%;  Score 20;  DB 9;  Length 15720;
Best Local Similarity   100.0%;  Pred. No. 5.1;
Matches   20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

Search completed: June 29, 2005, 21:23:40
Job time : 763 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: June 29, 2005, 17:01:40 ; Search time 4886 Seconds
                   (without alignments)
                   10294.006 Million cell updates/sec

Title: US-10-617-443B-1_COPY_1_1038
Perfect score: 1038
Sequence: 1 caggggactcgaagttgcag.........ggccccttccgccagcgcag 1038
Scoring table: Gapop 60.0 , Gapext 60.0
                   OLIGO-NUC
Searched: 4708233 seqs, 2422607955 residues
Word-size-1   12
Total number of hits satisfying chosen parameters: 730606
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :   GenEmbl:*
                   1: gb_ba:*
                   2: gb_htg:*
                   3: gb_in:*
                   4: gb_om:*
                   5: gb_ov:*
                   6: gb_pat:*
                   7: gb_ph:*
                   8: gb_pl:*
                   9: gb_pr:*
                   10: gb_ro:*
                   11: gb_sts:*
                   12: gb_sy:*
                   13: gb_un:*
                   14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	834	80.3	60597	9	AC113611	AC113611 Homo sapi
2	26	2.5	235102	2	AC118993	AC118993 Rattus no
3	26	2.5	258368	2	AC110369	AC110369 Rattus no
4	26	2.5	264075	2	AC134075	AC134075 Rattus no
5	24	2.3	145023	9	AC012621	AC012621 Homo sapi
6	24	2.3	164564	9	HUAC003108	HUAC003108 Human Chr
7	22	2.1	86027	9	AC137767	AC137767 Homo sapi
8	22	2.1	91841	9	AL831755	AL831755 Human DNA
9	22	2.1	97559	9	AL356748	AL356748 Human DNA
10	22	2.1	106497	9	AL157934	AL157934 Human DNA
11	22	2.1	112453	5	CQ869788	CQ869788 Sequence
12	22	2.1	125836	6	AC137629	AC137629 Homo sapi
13	22	2.1	128440	2	AC002419	AC002419 Homo sapi
14	22	2.1	131184	10	AL844513	AL844513 Mouse DNA
15	22	2.1	142118	9	AC104081	AC104081 Homo sapi
16	22	2.1	180396	9	AL445683	AL445683 Human DNA
17	22	2.1	183085	9	AC005815	AC005815 Homo sapi
18	22	2.1	192391	9	AC010768	AC010768 Homo sapi
19	22	2.1	193390	10	AC101527	AC101527 Mus muscu
20	22	2.1	197549	2	AC145423	AC145423 Homo sapi
21	22	2.1	220218	9	AC006288	AC006288 Homo sapi
22	22	2.1	233212	10	AL845161	AL845161 Mouse DNA
23	22	2.1	233874	2	BX842698	BX842698 Mus muscu
24	22	2.1	237538	2	AC108561	AC108561 Rattus no
25	22	2.1	243195	2	AC094523	AC094523 Rattus no
26	22	2.1	251121	2	AC128818	AC128818 Rattus no
27	21	2.0	1377	10	AF172275	AF172275 Mus muscu
28	21	2.0	1680	10	BC083769	BC083769 Rattus no
29	21	2.0	1795	10	AX740556	AX740556 Sequence
30	21	2.0	2045	10	BC025645	BC025645 Mus muscu
31	21	2.0	2055	6	CQ730199	CQ730199 Sequence
32	21	2.0	2184	9	CR456444	CR456444 Homo sapi
33	21	2.0	3088	9	AK025714	AK025714 Homo sapi
34	21	2.0	3200	9	BC015726	BC015726 Homo sapi
35	21	2.0	3236	9	BC040351	BC040351 Homo sapi
36	21	2.0	3944	9	BC027116	BC027116 Homo sapi
37	21	2.0	6112	10	AF075576	AF075576 Mus muscu
38	21	2.0	13178	9	AY504960	AY504960 Homo sapi
39	21	2.0	14268	10	AF338323	AF338323 Mus muscu
40	21	2.0	87077	9	HUAC004097	HUAC004097 Human Chr
41	21	2.0	91639	9	AL162399	AL162399 Human DNA
42	21	2.0	95283	9	HS599F21	HS599F21 Human DNA
43	21	2.0	110000	2	AL035662	AL035662 Human DNA
44	21	2.0	112071	4	BX649374_1	BX649374 Sinthops Continuation (2 of
45	21	2.0	120873	9	AL445986	AL445986 Human DNA

ALIGNMENTS

RESULT 1
AC113611
AC113611
```
LOCUS       AC113611               60597 bp    DNA     linear   PRI 10-MAR-2003
DEFINITION  Homo sapiens BAC clone RP11-421M20 from 4, complete sequence.
ACCESSION   AC113611
VERSION     AC113611.3  GI:28475761
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60597)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   MEDLINE   99063792
   PUBMED    9847074
REFERENCE   2  (bases 1 to 60597)
  AUTHORS   Isak,A. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-421M20
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 60597)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 60597)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 60597)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 60597)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
```

JOURNAL   Submitted (10-MAR-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Feb 23, 2003 this sequence version replaced gi:20136967.

COMMENT   ------- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@watson.wustl.edu
           -------- Summary Statistics
           Center project name: H_NH0421M20
           --------.

NOTICE: This sequence may not represent the entire insert of this
           clone. It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded, sequenced with an alternate
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

MAPPING INFORMATION:
           Mapping information for this clone was provided by Dr. John D.
           McPherson, Department of Genetics, Washington University, St. Louis
           MO. For additional information about the map position of this
           sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
           The RPCI-11 human BAC library was made from the blood of one male
           donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
           Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
           approach for construction of bacterial artificial chromosome
           libraries. Genomics 51:1-8. The clone may be obtained either from
           Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
           and coworkers at http://www.chori.org
           VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
           This sequence is not the entire insert of the clone. This clone is
           overlapped by AC104650 and AC104825.

           Data from AC104825 was used to finish this clone.

           Discrepant bases between AC104650, AC104825 and clone sequence.

FEATURES   Location/Qualifiers
     source   1..60597
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-421M20"
                     /clone_lib="RPCI-11"
     repeat_region   539..903
                     /rpt_family="ERVL"
     repeat_region   2797..2907
                     /rpt_family="(CAGAGA)n"
     repeat_region   2908..3114
                     /rpt_family="ERV1"
     repeat_region   3122..3184
                     /rpt_family="(CAGAGA)n"
     repeat_region   3256..3368
                     /rpt_family="L1"
     repeat_region   3500..3682
                     /rpt_family="L1"
     repeat_region   3713..3782
                     /rpt_family="AT_rich"
     repeat_region   5208..5410
                     /rpt_family="MIR"

     repeat_region   5451..5503
                     /rpt_family="MIR"
     repeat_region   5504..5677
                     /rpt_family="MIR"
     repeat_region   6645..6687
                     /rpt_family="MIR"
     repeat_region   6998..7067
                     /rpt_family="MIR"
     repeat_region   7741..8173
                     /rpt_family="MIR"
     repeat_region   8530..8610
                     /rpt_family="MER1_type"
     repeat_region   8638..9514
                     /rpt_family="MIR"
     misc_feature   9027..9048
                     /note="CpG_island (%GC=76.7, o/e=0.80, #CpGs=114)"
     repeat_region   9063..9111
                     /rpt_family="GC_rich"
     repeat_region   9612..9739
                     /rpt_family="GC_rich"
     repeat_region   10414..10509
                     /rpt_family="MIR"
     repeat_region   10688..11034
                     /rpt_family="MER103"
     repeat_region   11781..11997
                     /rpt_family="MaLR"
     repeat_region   12212..12325
                     /rpt_family="L2"
     repeat_region   13275..13625
                     /rpt_family="MIR"
     repeat_region   13703..13769
                     /rpt_family="L2"
     repeat_region   14926..15054
                     /rpt_family="CR1"
     repeat_region   15239..15281
                     /rpt_family="MIR"
     repeat_region   15349..15515
                     /rpt_family="Tip100"
     repeat_region   16000..16613
                     /rpt_family="MIR"
     repeat_region   16639..16917
                     /rpt_family="MIR"
     repeat_region   17127..17228
                     /rpt_family="L2"
     repeat_region   17883..17982
                     /rpt_family="Tip100"
     repeat_region   18108..18173
                     /rpt_family="MIR"
     repeat_region   19059..19127
                     /rpt_family="CT-rich"
     repeat_region   20087..20234
                     /rpt_family="L2"
     repeat_region   20576..20738
                     /rpt_family="MIR"
     repeat_region   22242..22380
                     /rpt_family="MIR"
     repeat_region   23733..23804
                     /rpt_family="(TGAA)n"
     repeat_region   23890..23987
                     /rpt_family="L2"
     repeat_region   24011..24440
                     /rpt_family="MIR"
     repeat_region   25254..25412
                     /rpt_family="MaLR"
     repeat_region   26001..26150
                     /rpt_family="MIR"
     repeat_region   26473..26512
                     /rpt_family="MIR"
     repeat_region   26513..26894
                     /rpt_family="MIR"
     repeat_region   26895..27048
                     /rpt_family="MaLR"
     repeat_region   27423..27724
                     /rpt_family="MIR"

```
repeat_region   27725..28016
                /rpt_family="Alu"
repeat_region   28029..28096
                /rpt_family="Alu"
repeat_region   28099..28399
                /rpt_family="(TTCC)n"
repeat_region   28406..28597
                /rpt_family="Alu"
repeat_region   28597..29783
                /rpt_family="MER1_type"
repeat_region   29613..29783
                /rpt_family="L1"
repeat_region   29988..30078
                /rpt_family="L1"
repeat_region   30801..30893
                /rpt_family="MIR"
```

```
Query Match    80.3%;  Score 834;  DB 9;  Length 60597;
Best Local Similarity  99.6%;  Pred. No. 0;
Matches 1034;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;
```

```
Qy       1 CAGGGACTCGAAGTTTGCAGTCCTCCTCCACACTCAGTTCCCACAGATGTGGGTAGGAGGGCAT     60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20549 CAGGGACTCGAAGTTTGCAGTCCTCCTCCACACTCAGTTCCCACAGATGTGGGTAGGAGGGCAT  20608

Qy      61 ATTCAGTCCCATTTTTCAGATGAGGAAGTTGAGGCCCAGAGGAACGTAAGTAATCTGTCTGA     120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20609 ATTCAGTCCCATTTTTCAGATGAGGAAGTTGAGGCCCAGAGGAACGTAAGTAATCTGTCTGA  20668

Qy     121 GGCCACACAGCTAGAAAAGCAGCCCAGGCCCGGAACCCCTGGGTTGTGCAGCCCCCAGC     180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20669 GGCCACACAGCTAGAAAAGCAGCCCAGGCCCGGAACCCCTGGGTTGTGCAGCCCCCAGC  20728

Qy     181 CCAGTTGCTCATTCGCGGGGGCTCGGGGACCCACGACGCCGAGGCCAGA     240
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   20729 CCAGTTGCTCATTCGCGGGGGCTCGGGGACCCACGACGCCGAGGCCAGA  20788

Qy     241 TGGTGGGAACTGGAGAGAGGCCCGGCCACAGCCCGTGCACGGGAACCCCGAGGGGCTGTAGGC     300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20789 TGGTGGGAACTGGAGAGAGGCCCGGCCACAGCCCGTGCACGGGAACCCCGAGGGGCTGTAGGC  20848

Qy     301 CCCGTGCCACTGCATGCCTCAGGGCCTCGTGTGGTCTCGGCACCCACCAGCCCCTTACTGCTGACG     360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20849 CCCGTGCCACTGCATGCCTCAGGGCCTCGTGTGGTCTCGGCACCCACCAGCCCCTTACTGCTGACG  20908

Qy     361 GCAGCAGGAGAATCTGAGCCCGGGGAAGGGTCCAGGGAAGGTTCGTGAACCATCTAGCAAGTCG     420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20909 GCAGCAGGAGAATCTGAGCCCGGGGAAGGGTCCAGGGAAGGTTCGTGAACCATCTAGCAAGTCG  20968

Qy     421 GGCTGGGGTGTGGCCAAGTTAGACACAGATGTAGGGCCCTGTGGACTCAGAAATTGGCAG     480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20969 GGCTGGGGTGTGGCCAAGTTAGACACAGATGTAGGGCCCTGTGGACTCAGAAATTGGCAG  21028

Qy     481 CTCTTTTTGCCCCAGAGGGCCGCCACACCCTGTGTCGTCGTGCGTAGCTCAGAAGGGTCACCT     540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21029 CTCTTTTTGCCCCAGAGGGCCGCCACACCCTGTGTCGTCGTGCGTAGCTCAGAAGGGTCACCT  21088

Qy     541 GGGGGTCTTCCACTACACCCCCGGGACACTGCGTGTAGCCTCGTGGAGGGACC     600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21089 GGGGGTCTTCCACTACACCCCCGGGACACTGCGTGTAGCCTCGTGGAGGGACC  21148

Qy     601 AGTGTGGAGCCCATGAGGAGGAGGCCAGTGTTCTCTCCTGTAAGGGGTATTGCTGTAGGATGAG     660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21149 AGTGTGGAGCCCATGAGGAGGAGGCCAGTGTTCTCTCCTGTAAGGGGTATTGCTGTAGGATGAG  21208

Qy     661 GGAACAGCAAAGGCCCAGAGGGGGACTAACCCGGAGATCCAGCCCTCACTCCCGTGTG     720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21209 GGAACAGCAAAGGCCCAGAGGGGGACTAACCCGGAGATCCAGCCCTCACTCCCGTGTG  21268

Qy     721 GCTCACGGCAATATCCTAACCTCTCTCTGAGCCTTCCTCGCCTAGCGAGGGTCCAGTG     780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21269 GCTCACGGCAATATCCTAACCTCTCTCTGAGCCTTCCTCGCCTAGCGAGGGTCCAGTG  21328

Qy     781 AGGGGGGTGAGGAAGCCCAGCACGTGGGAAGCCCTTTTTAACCATTCTCGGGGGTTGAGCCGAGC     840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21329 AGGGGGGTGAGGAAGCCCAGCACGTGGGAAGCCCTTTTTAACCATTCTCGGGGGTTGAGCCGAGC  21388

Qy     841 CCCTTCCCAAATGCCGTGGTGTCACTGCATGCTGTGTGGTGGTAGGGGGGTGTCCCCAACGGGCTC     900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21389 CCCTTCCCAAATGCCGTGGTGTCACTGCATGCTGTGTGGTGGTAGGGGGGTGTCCCCAACGGGCTC  21448

Qy     901 AGTGTGGGGTCGGCTGGAGGCGTGGCTCTGAACTGGGACAGGGGGTCTCAGGAGAGAGCCTCCTCC     960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21449 AGTGTGGGGTCGGCTGGAGGCGTGGCTCTGAACTGGGACAGGGGGTCTCAGGAGAGAGCCTCCTCC  21508

Qy     961 TGCCCACTGGGGCATAGGGCCCTCTGGCCGAGCCGAGCATGGGTCTCAGGAAGAGAGCCCTTC    1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21509 TGCCCACTGGGGCATAGGGCCCTCTGGCCGAGCCGAGCATGGGTCTCAGGAAGAGAGCCCTTC  21568

Qy    1021 CCCTTCCCGCGCCAGCGCGCAG    1038
           ||||||||||||||||||||||
Db   21569 CCCTTCCCGCGCCAGCGCGCAG  21586
```

```
RESULT 2
AC118993/c
LOCUS       AC118993    235102 bp    DNA    linear   HTG  15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-237H8, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC118993
VERSION     AC118993.4  GI:25013258
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 235102)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Derano,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Forster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Liu,Y., London,P., Longacre,S., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunnaagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Reeves,K., Pu,L., L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
```

```
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 235102)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (24-APR-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
REFERENCE  3  (bases 1 to 235102)
AUTHORS    Rat Genome Sequencing Consortium.
TITLE      Direct Submission
JOURNAL    Direct Submission
           Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Nov 15, 2002 this sequence version replaced gi:23681105.
           The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each 'contig-scaffold,
           individual sequence contigs are ordered and oriented, and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the ends of the clone and there may be multiple
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contigs will be indicated in the feature
           table.
           ---------------  Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu/
           Contact: hgsc-help@bcm.tmc.edu
           ---------------  Project Information
           Center project name: GTNH
           Center clone name: CH230-237H8
           ---------------  Summary Statistics
           Assembly program: Phrap; version 0.990329
           Consensus quality:  227037 bases at least Q40
           Consensus quality:  229537 bases at least Q30
           Consensus quality:  231225 bases at least Q20
           Consensus quality:  235071; sum-of-contigs estimation
           Estimated insert size: 235071; sum-of-contigs estimation
           Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
           ------------------------------------------------------
           NOTE: Estimated insert size may differ from sequence length
             (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
           NOTE: This is a 'working draft' sequence. It currently
           consists of 2 contigs. The true order of the pieces
           is not known and their order in this sequence record is
           arbitrary. Gaps between the contigs are represented as
           runs of N, but the exact sizes of the gaps are unknown.
           This record will be updated with the finished sequence
           as soon as it is available and the accession number will
           be preserved.
           ------------------------------------------------------
           *    108068: contig of 108068 bp in length
           *    108069: gap of unknown length
           *    108069  108168: gap of unknown length
           *    108169  235102: contig of 126934 bp in length.
FEATURES             Location/Qualifiers
     source          1..235102
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-237H8"
     misc_feature    1..1143
                     /note="wgs_end_extension"
```
```
     misc_feature    complement(3003..3706)
                     /note="clone_boundary"
                     /clone_end:Sp6
                     site:
                     end_sequence:B2101611"
                     108169..110729
     misc_feature    108169..110729
                     /note="wgs_contig"

ORIGIN

Qy       313 CATGCCTCAGGCCTGTGGTCCTGGCA 338
             |||||||||||||||||||||||||||
Db     89157 CATGCCTCAGGCCTGTGGTCCTGGCA 89132

RESULT 3
AC110369//c
LOCUS       AC110369       258368 bp    DNA    linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-49K10, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC110369
VERSION     AC110369.4  GI:23195246
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258368)
  AUTHORS   Muzny,D.Marie., Metzker,M.lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,A.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,J., Jiang,H., Johnson,B., Johnson,D., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahidantne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,R., Martin,N., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,P., Popovic,D., Primus,E.P., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,E., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
```

Qy   2.5%; Score 26; DB 2; Length 235102;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db   89157 CATGCCTCAGGCCTGTGGTCCTGGCA 89132

(bases 1 to 235102)
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
```

```
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,R., Warren,R., Wei,X., White,F.,
            Williams,G., Wilson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 258368)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine,
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 258368)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2002 this sequence version replaced gi:21744398.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads. Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            --------------------------------------
                          Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------------------------------
                          Project Information
            Center project name: GRPY
            Center clone name: CH230-49K10
            --------------------------------------
                          Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 23638 bases at least Q40
            Consensus quality: 23638 bases at least Q40
            Consensus quality: 238877 bases at least Q30
            Consensus quality: 24071 bases at least Q20
            Estimated insert size: 259061; sum-of-contigs estimation
            Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------------------------------
            NOTE: Estimated insert size may differ from sequence length
            (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            NOTE: This sequence may represent more than one clone.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 3 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *   1               102159: contig of 102159 bp in length
            *   102159          102259: gap of unknown length
            *   102160          102259: gap of unknown length
            *   102260          249257: contig of 146998 bp in length
            *   249258          249357: gap of unknown length
            *   249358          258368: gap of unknown length
            *   249358          258368: contig of 9011 bp in length.
FEATURES             Location/Qualifiers
     source          1..258368
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-49K10"
     misc_feature    1..1248
                     /note="wgs end extension
                     /clone_end:T7"
     misc_feature    6211..7032
```

```
                     /note="clone boundary
                     clone_end:T7"
                     site:EcoRI
                     end sequence:BH282841"
     misc_feature    78223..79518
                     /note="wgs_contig"
     misc_feature    102260..104120
                     /note="wgs_contig"

ORIGIN
     Query Match             2.5%;  Score 26;  DB 2;  Length 258368;
     Best Local Similarity 100.0%;  Pred. No. 0.0057;
     Matches 26;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy        313 CATGCCTCAGGCCTGTGGTCCTGGCA 338
              ||||||||||||||||||||||||||
Db      82091 CATGCCTCAGGCCTGTGGTCCTGGCA 82066

RESULT 4
AC134075
LOCUS       AC134075              264075 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2H18, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC134075
VERSION     AC134075.2  GI:24941374
KEYWORDS    HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 264075)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kurehi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L-L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,J., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
```

TITLE
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 264075)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 264075)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 13, 2002 this sequence version replaced gi:23267440.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
--------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
--------------- Project Information
Center project name: TUQB
Center clone name: CH230-2H18
--------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 225336 bases at least Q40
Consensus quality: 228408 bases at least Q30
Consensus quality: 230612 bases at least Q20
Estimated insert size: 229980; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft.html).
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 2 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
FEATURES Location/Qualifiers
source 1..264075
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-2H18"
misc_feature 1 262529: contig of 262529 bp in length
misc_feature 262530 262629: gap of unknown length
misc_feature 262630 264075: contig of 1446 bp in length.
misc_feature
clone_end:T7"
misc_feature 1858..153329
/note="clone_boundary
clone_end:T7"
misc_feature 1..1475
/note="wgs_end_extension
site:
end_sequence:BH287375"
site:
258169..258875
/note="clone_boundary
clone_end:Sp6"

misc_feature
site:
end_sequence:BH287377"
259534..262529
/note="wgs_end_extension
clone_end:Sp6"
ORIGIN
Qy   313 CATGCCTCAGGCCTGTGGGTCCTGGCA 338
          |||||||||||||||||||||||||||
Db 88708 CATGCCTCAGGCCTGTGGGTCCTGGCA 88733

Query Match 2.5%; Score 26; DB 2; Length 264075;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
AC012621
LOCUS   AC012621 145023 bp DNA linear PRI 18-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone CTD-2055J8, complete sequence.
ACCESSION   AC012621
VERSION   AC012621.8 GI:29029223
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 145023)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 145023)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 145023)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 145023)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
TITLE   Direct Submission
JOURNAL   Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Mar 18, 2003 this sequence version replaced gi:18542968.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES Location/Qualifiers
source 1..145023
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2055J8"
ORIGIN
Qy   118 TGAGGCCACACAGCTAGAAAAGCAG 141
          |||||||||||||||||||||||||
Db 99318 TGAGGCCACACAGCTAGAAAAGCAG 99341

Query Match 2.3%; Score 24; DB 9; Length 145023;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
RESULT 6
LOCUS         HUAC003108   164564 bp    DNA    linear   PRI 30-OCT-2002
DEFINITION    Human Chromosome 16 BAC clone CIT987SK-327024; complete sequence.
ACCESSION     AC003108
VERSION       AC003108.1  GI:2833632
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 164564)
AUTHORS       Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
              Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
              Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
              Eichler,E.E.,Harris,P.C., Venter,J.C. and Adams,M.D.
TITLE         Genome duplications and other features in 12 Mb of DNA sequence
              from human chromosome 16p and 16q
JOURNAL       Genomics 60 (3), 295-308 (1999)
MEDLINE       99452270
PUBMED        10493829
REFERENCE     2  (bases 1 to 164564)
AUTHORS       Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
              Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
TITLE         Human Chromosome 16 BAC clone CIT987SK-327024
JOURNAL       Unpublished
REFERENCE     3  (bases 1 to 164564)
AUTHORS       Adams,M.D. and Loftus,B.J.
TITLE         Direct Submission
JOURNAL       Submitted (19-NOV-1997) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA, Email:
              bjloftus@tigr.org
REFERENCE     4  (bases 1 to 164564)
AUTHORS       Adams,M.D.
TITLE         Direct Submission
JOURNAL       Submitted (05-FEB-1998) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA
REFERENCE     5  (bases 1 to 164564)
AUTHORS       Adams,M.D. and Loftus,B.J.
TITLE         Direct Submission
JOURNAL       Submitted (17-MAR-1998) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA
COMMENT       On Feb 5, 1998 this sequence version replaced gi:2827780.
              Address all correspondence to: Mark Adams The Institute for Genomic
              Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
              address: mdadams@tigr.org. The orientation of the sequence is from
              SP6 end to T7 end. Genes were identified by a combination of five
              methods including: XGRAIL (available by anonymous ftp from
              arthur.epm.ornl.gov), Genefinder (Phil Green, University of
              Washington), Genscan (Chris Burge,
              http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
              complete sequence against a peptide database,and the Human gene
              Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
              Genes without peptide homology having spliced EST hits are termed
              'Unknown gene product'. Genes encoding tRNAs are predicted by
              tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1..164564
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p11.2-12"
                     /clone="327024"
     STS             9637..9967
                     /db_xref="dbSTS:G21493"
     gene            complement(10088..28580)
                     /gene="327024_1"
                     /note="327024_1"
     mRNA            complement(join(10088..10235,11499..11710,14305..14397,
                     17457..17562,23632..23807,28068..28253,28320..28580))
                     /gene="327024_1"
```

```
     CDS             complement(join(10088..10235,11499..11710,14305..14397,
                     17457..17562,23632..23807,28068..28253,28320..28580))
                     /gene="327024_1"
                     /codon_start=1
                     /product="Unknown gene product"
                     /protein_id="AAC05803.1"
                     /db_xref="GI:2833633"
                     /translation="MWLWEDQGGLLGPFSFLLLVLLVTRSPVNACLLLTQSSLFVLLRV
                     FSFEFVPSCRALQVLKPDRISAIAHRGGSHDAPENTLAAIRQGSPIFSGRPSHELIR
                     PQGTLPALDRSQCLGISGLEVLYLQKTADMYIVSVRLITLRHTARNKAARNGATGVEL
                     DIEFTSDGIPVLMHDNTVDRTTDGTGRLCDLITFEGIRKLNPAANHLRLRNDFPDEKIPT
                     LREAVAECLNHNLTIFPDVKGHAHKATEALKKMMEFPQLYNNSVVCSFLPEVIYKMR
                     QTDRDVITALTHRPWSLSHTGDGKRYDTFWKHFIFVVMDILLDWSMHNILWYLCGIS
                     AFLMQKDFVSPAYLKKWWSAKGIQVVGWTVNTFDEKSYYESHLGGSSYITDSMVEDCEPH
                     F"
     misc_feature    28025..29390
                     /note="presence of CpG island"
     gene            32335..60015
                     /gene="327024_2"
     mRNA            join(32335..32477,34498,34638,39032,39160,42555,44202,
                     47273..47403,48499..48676,49233..49372,49490..49606,
                     51409..51510,51706..51822,52429..52502,53024..53086,
                     54567..54626,57808..57883)
                     /gene="327024_2"
     CDS             join(34498,34638,39032,39160,42555,44202,47273,47403,
                     48499..48676,49233..49372,49490..49606,51409..51510,
                     51706..51822,52429..52502,53024..53086,54567..54626,
                     57808..57883)
                     /gene="327024_2"
                     /codon_start=1
                     /product="KIAA0420"
                     /protein_id="AAC05804.1"
                     /db_xref="GI:2833634"
                     /translation="MEEYEKFCEKSLARIQEASLSTESFLPAQSESISILRFHGVAIL
                     SPLINIEKRKEMQQEKQKALDVEARKQVNRKKALLTRVQSILDNVQVRKAPNASDFDQ
                     WEMETVYYSNSEVRNILNVPATFPNSPSHTEHSTRAPKLDKIAGILPLDNEDQCKTDGID
                     LARDSEGFPNSPKQCDSSNISHVENEAFPKTSSATPQEETLISDGPSVNEQQDLPLLAE
                     VIPDPYVMSLQNLMKKSKEYIEREQSRSLRGSMNRIVNESHLDKEHDAVEVADCVKE
                     KGQLTGKHCVSVIPDKPSLNKSNVTLLQGASTQASSMSMPVLASFSKVDIPIRTGHPTV
                     LESNSDFKVITPFVTVENNVIKSLTGSSYAKLPSEPSMSPKGHRRRSRTSSACHILINN
                     PINACELSPKGKEQAMDLIIQDTDENTNVPEIMPKLPTDLAGVCSSKVVGKNTSEVK
                     EDVVLGKSNQVCQSSGNHIENKVTHGLVTVEGQITSDERGAHIMNSTCAAMPKLHEPY
                     ASSQCIASNFGTVSGLKPASMEKNCSLQTELNKSYDVKNTEQPYVITSGITEQERQHLPEKRYPKG
                     PMVSCGNEQPLDNSFEKVKRRLDIDIGLDKCKENCPYVITSGITEQEROHLPEKRYPKG
                     SGFVNNKMLGTSSKSEEBLLKSKNLAFEEMRKRLEEQHAQQLSLLIAEQEREQRLQ
                     KEIEEQKMLKEKAVMTAEASELDINNAVELEWRKISDSSLLETMLSQADSLIHTSNSN
                     SSGFTNSAMQYSFVSANEAPFYLWGSSTSGLTKLSVTRPGRAKTRWSQVFSLEIQAK
                     FNKITAVAKGFLTRLRLMQTDKLKQLRQTVDTMEFIRSFQSEAPLKRGIVSAQDASLQ
                     ERVLAQLRAALYGIHDIFPVMDAAERMSILHHDEVRKEKMLRQMDKMKSPRVALSAA
                     TQKSLDRKXYMKAAEMGMPNKKFLVKQNPSETRVLQPNQQONAPVHRLLSRQGSICRK
                     NPKAAKCCDNLRQHSLG"
ORIGIN
Qy        118 TGAGGCCACACAGCTAGAAAGCAG 141
              ||||||||||||||||||||||||
Db     123634 TGAGGCCACACAGCTAGAAAGCAG 123657
```

```
        Query Match          2.3%;  Score 24;  DB 9;  Length 164564;
        Best Local Similarity  100.0%;  Pred. No. 0.084;
        Matches  24;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT 7
AC137767
LOCUS         AC137767            86027 bp    DNA    linear   PRI 21-JUN-2003
DEFINITION    Homo sapiens 12 BAC RP13-942N8 (Roswell Park Cancer Institute Human
              BAC Library) complete sequence.
ACCESSION     AC137767
VERSION       AC137767.16  GI:32140199
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

  ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 86027)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Binage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Caroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,J., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Louleeged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Mattindale,A., Mattinez,E., Massey,E.,
            Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,Z., Miner,G., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newton,J., Newson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peter6,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 86027)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-DEC-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 86027)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 86027)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-JUN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5 (bases 1 to 86027)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-JUN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jun 21, 2003 this sequence version replaced gi:3137640.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clone.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.
FEATURES   Location/Qualifiers
     source   1..86027
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP13-942N8"
     misc_feature   1..2000
                     /note="overlaps bases 116317..118312 of clone AC055713"
                     /function="clone overlap"
     repeat_region   complement(1..238)
                     /rpt_family="AluJb"
     repeat_region   complement(252..549)
                     /rpt_family="AluSg"
     repeat_region   682..992
                     /rpt_family="AluSg"
     repeat_region   1180..1472
                     /rpt_family="AluY"
     repeat_region   1483..1780
                     /rpt_family="AluSx"
     repeat_region   1826..2101
                     /rpt_family="AluSx"
     repeat_region   complement(2505..2660)
                     /rpt_family="LiMEl"
     repeat_region   complement(2669..2831)
                     /rpt_family="MIRb"
     repeat_region   3186..3415
                     /rpt_family="LiMB5"
     repeat_region   3966..4079
                     /rpt_family="Charlie8"
     repeat_region   4092..4384
                     /rpt_family="AluSg"
     repeat_region   complement(4858..4944)
                     /rpt_family="L2"
     repeat_region   4941..5110
                     /rpt_family="L2"
     repeat_region   5113..5409
                     /rpt_family="AluJo"

RESULT 8

```
Qy    114 TGTCTCGAGGGCCACACAGCCTAGA 135
          |||||||||||||||||||||||||
Db   2755 TGTCTCGAGGGCCACACAGCCTAGA 2776
```

Query Match   2.1%; Score 22; DB 9; Length 86027;
Best Local Similarity   100.0%;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
repeat_region   5413..5720
                /rpt_family="AluSx"
repeat_region   5757..5857
                /rpt_family="L1PA15-16"
repeat_region   5865..6155
                /rpt_family="AluY"
repeat_region   6200..6490
                /rpt_family="AluSg"
repeat_region   6492..6635
                /rpt_family="AluSx"
repeat_region   6636..6942
                /rpt_family="L1MD3"
repeat_region   complement(6943..6957)
                /rpt_family="AluSg"
repeat_region   6960..7140
                /rpt_family="L1MD3"
repeat_region   7173..7475
                /rpt_family="AluY"
repeat_region   7569..7676
                /rpt_family="L2"
repeat_region   complement(7482..7568)
                /rpt_family="FLAM_A"
repeat_region   complement(8998..9293)
                /rpt_family="AluSx"
repeat_region   9322..9630
                /rpt_family="AluSp"
repeat_region   9752..9773
                /rpt_family="AT_rich"
repeat_region   complement(10253..10437)
                /rpt_family="MIR"
repeat_region   11149..11425
                /rpt_family="AluJb"
repeat_region   11432..11734
                /rpt_family="AluSg"
repeat_region   11759..11779
                /rpt_family="AluSg"
repeat_region   12892..13199
                /rpt_family="AT_rich"
repeat_region   13226..13505
                /rpt_family="AluSx"
repeat_region   13506..13754
                /rpt_family="L1MB7"
repeat_region   13755..14089
                /rpt_family="AluSx"
repeat_region   14418..14417
                /rpt_family="L1MB7"
```

LOCUS   AL831755/c   91841 bp   DNA   linear   PRI 26-SEP-2002
DEFINITION Human DNA sequence from clone RP13-279N23 on chromosome 1, complete
            sequence.
ACCESSION   AL831755
VERSION   AL831755.9 GI:23337613
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 91841)
AUTHORS   Harrison,E.
TITLE   Direct Submission
JOURNAL   Submitted (26-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

---

COMMENT   On Sep 27, 2002 this sequence version replaced gi:22798452.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk

            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP13-279N23 is from the library RPCI-13.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES   Location/Qualifiers
     source   1..91841
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP13-279N23"
                     /clone_lib="RPCI-13.2"
ORIGIN

Query Match   2.1%; Score 22; DB 9; Length 91841;
Best Local Similarity   100.0%;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    113 CTGTCTCGAGGGCCACACAGCTAG 134
          ||||||||||||||||||||||||
Db  83423 CTGTCTCGAGGGCCACACAGCTAG 83402
```

RESULT 9

LOCUS   AL356748   97559 bp   DNA   linear   PRI 04-DEC-2000
DEFINITION Human DNA sequence from clone RP4-686G16 on chromosome 1, complete
            sequence.
ACCESSION   AL356748
VERSION   AL356748.20 GI:11558567
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 97559)
AUTHORS   Laird,G.
TITLE   Direct Submission
JOURNAL   Submitted (03-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT   On Dec 5, 2000 this sequence version replaced gi:11493283.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP4-686J16 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RP4-686J16 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP4-686J16 is at 1 in this sequence. The true left end of clone RP4-655N15 is at 97460 in this sequence. The true right end of clone RP4-781D12 is at 1573 in this sequence.

```
FEATURES             Location/Qualifiers
source               1..97559
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP4-686J16"
                     /clone_lib="RPCI-4"
repeat_region        378..514
                     /note="MIR repeat: matches 81..226 of consensus"
repeat_region        927..1084
                     /note="MIR repeat: matches 100..258 of consensus"
repeat_region        1112..1353
                     /note="AluSx repeat: matches 1..256 of consensus"
repeat_region        1512..1812
                     /note="AluSx repeat: matches 1..300 of consensus"
misc_feature         complement(1745..2002)
                     /note="match: GSS: Em:AQ321132"
repeat_region        1820..1927
                     /note="AluJb repeat: matches 191..298 of consensus"
misc_feature         2008..2639
                     /note="match: GSS: Em:AQ474986"
misc_feature         2014..2416
                     /note="match: GSS: Em:AQ706294"
repeat_region        2051..2205
                     /note="AluSx repeat: matches 1..256 of consensus"
repeat_region        complement(2328..2414)
                     /note="MIR repeat: matches 78..252 of consensus"
misc_feature         complement(2908..3284)
                     /note="MLT1F repeat: matches 401..534 of consensus"
repeat_region        2477..2555
                     /note="MLT1H repeat: matches 51..130 of consensus"
repeat_region        2676..2813
                     /note="MLT1H repeat: matches 202..288 of consensus"
misc_feature         complement(3270..3365)
                     /note="match: STS: Em:G18185"
repeat_region        4118..4212
                     /note="L1P3 repeat: matches 5725..5822 of consensus"
repeat_region        7092..7341
                     /note="L2 repeat: matches 2075..2178 of consensus"
misc_feature         complement(7568..8088)
                     /note="match: STS: Em:G15587"
misc_feature         complement(7909..8281)
                     /note="match: GSS: Em:AQ078880"
misc_feature         8073..8228
                     /note="match: GSS: Em:AQ035467"

repeat_region        8286..8342
                     /note="MIR repeat: matches 106..245 of consensus"
repeat_region        8660..8744
                     /note="L2 repeat: matches 2691..2749 of consensus"
repeat_region        8745..9049
                     /note="L2 repeat: matches 2217..2298 of consensus"
repeat_region        9050..9230
                     /note="AluY repeat: matches 1..305 of consensus"
repeat_region        complement(10541..11084)
                     /note="L2 repeat: matches 2023..2217 of consensus"
misc_feature         complement(10633..10966)
                     /note="match: GSS: Em:AQ316052"
misc_feature         11304..11419
                     /note="match: GSS: Em:AZ121092"
repeat_region        12397..12857
                     /note="LTR9B repeat: matches 445..560 of consensus"
misc_feature         complement(13076..13897)
                     /note="MLT1G repeat: matches 31..512 of consensus"
repeat_region        13492..13591
                     /note="match: GSS: Em:AQ743494"
repeat_region        13685..13725
                     /note="AluSx repeat: matches 5..312 of consensus"
repeat_region        14198..14273
                     /note="L2 repeat: matches 2388..2487 of consensus"
repeat_region        15025..15148
                     /note="L2 repeat: matches 2709..2749 of consensus"
repeat_region        16477..16612
                     /note="L2 repeat: matches 2673..2749 of consensus"
misc_feature         complement(16832..17296)
                     /note="L2 repeat: matches 2584..2708 of consensus"
repeat_region        17109..17167
                     /note="MIR repeat: matches 85..228 of consensus"
misc_feature         17344..17992
                     /note="match: GSS: Em:AQ700892"
repeat_region        17496..17773
                     /note="MER45B repeat: matches 437..492 of consensus"
repeat_region        18470..18669
                     /note="match: GSS: Em:B50102"
repeat_region        18751..18902
                     /note="AluSq repeat: matches 37..311 of consensus"
repeat_region        18928..19233
                     /note="L1MC/D repeat: matches 5423..5612 of consensus"
repeat_region        19235..19296
                     /note="L1MC5 repeat: matches 7758..7913 of consensus"
repeat_region        complement(19553..19954)
                     /note="AluSx repeat: matches 5..312 of consensus"
repeat_region        complement(19692..19772)
                     /note="31 copies 2 mer aa 74% conserved"
repeat_region        20881..20991
                     /note="Alu7/FLAM repeat: matches 5..85 of consensus"
repeat_region        21460..21617
                     /note="AluJo/FRAM repeat: matches 203..304 of consensus"
repeat_region        complement(22616..23224)
                     /note="MIR repeat: matches 16..142 of consensus"
repeat_region        23288..23435
                     /note="L1ME1 repeat: matches 5528..6159 of consensus"
repeat_region        complement(25186..25354)
                     /note="MIR repeat: matches 20..184 of consensus"
repeat_region        25530..25827
                     /note="MIR repeat: matches 80..233 of consensus"
misc_feature         complement(26592..27148)
                     /note="MIR repeat: matches 77..262 of consensus"
misc_feature         complement(26598..27124)
                     /note="AluSx repeat: matches 1..298 of consensus"
misc_feature         complement(26720..27017)
                     /note="match: GSS: Em:AQ532789"
misc_feature         complement(26833..27131)
                     /note="match: GSS: Em:AQ619294"
repeat_region        26971..27041
                     /note="match: GSS: Em:AQ100816"
misc_feature         27111..27147
                     /note="match: GSS: Em:AQ837606"
repeat_region        27111..27147
                     /note="MLT1J repeat: matches 110..180 of consensus"
misc_feature
                     /note="Single clone region. Assembly confirmed by"
```

Left column

```
misc_feature    27131..27224
                /note="match: GSS: Em:AQ136115"
misc_feature    27133..27439
                /note="match: GSS: Em:AQ556313"
misc_feature    27119..27631
                /note="match: GSS: Em:AQ798882"
repeat_region   27202..27252
                /note="3 copies 17 mer 86% conserved"
misc_feature    27218..27578
                /note="match: GSS: Em:AQ438628"
misc_feature    27236..27568
                /note="match: GSS: Em:AQ136115"
repeat_region   29394..29510
                /note="match: GSS: Em:AQ316053"
repeat_region   29510
                /note="AluY repeat: matches 17..311 of consensus"
repeat_region   29562..29638
                /note="L2 repeat: matches 2587..2699 of consensus"
repeat_region   32900..32937
                /note="AluSx repeat: matches 1..287 of consensus"
repeat_region   33091..33356
                /note="L2 repeat: matches 2660..2705 of consensus"
misc_feature    33999..34281
                /note="match: GSS: Em:AQ316053"
repeat_region   33435..34640
                /note="AluY repeat: matches 17..311 of consensus"
repeat_region   35737..36047
                /note="AluSx repeat: matches 1..287 of consensus"
repeat_region   37307..37627
                /note="AluSq repeat: matches 1..313 of consensus"
repeat_region   38057..38173
                /note="AluJo repeat: matches 1..293 of consensus"
repeat_region   38174..38190
                /note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region   38191..38449
                /note="Charlie4 repeat: matches 1855..1871 of consensus"
                /note="AluSx repeat: matches 1..302 of consensus"
```

```
Qy          113 CTGTCTGAGGCCACACAGCTAG 134
                ||||||||||||||||||||||
Db        20910 CTGTCTGAGGCCACACAGCTAG 20931
```

Query Match   2.1%; Score 22; DB 9; Length 97559;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
LOCUS   AL157934   106497 bp   DNA   linear   PRI 06-JAN-2002
AL157934/c
DEFINITION   Human DNA sequence from clone RP11-449M9 on chromosome Xq13.1-13.3,
             complete sequence.
ACCESSION   AL157934
VERSION   AL157934.17 GI:18077665
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 106497)
AUTHORS   Wray,P.
TITLE   Direct Submission
JOURNAL   Submitted (05-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Jan 6, 2002 this sequence version replaced gi:17381304.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence was finished as follows unless otherwise noted: all
             regions were either double-stranded or sequenced with an alternate
             chemistry or covered by high quality data (i.e., phred quality >=
```

## Right column

```
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by at least
             one plasmid subclone or more than one M13 subclone; and the
             assembly was confirmed by restriction digest. The following
             abbreviations are used to associate primary accession numbers given
             in the feature table with their source databases: Em:, EMBL; Sw:,
             SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information on the WORMPEP
             database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
             was generated from part of bacterial clone contigs of human
             chromosome X, constructed by the Sanger Centre Chromosome X Mapping
             Group. Further information can be found at
             http://www.sanger.ac.uk/HGP/ChrX
             RP11-449M9 is from the library RPCI-11.2 constructed by the group
             of Pieter de Jong. For further details see
             http://www.chori.org/bacpac/home.htm
             VECTOR: pBACe3.6
             IMPORTANT: This sequence is not the entire insert of clone
             RP11-449M9 It may be shorter because we sequence overlapping
             sections only once, except for a short overlap.
             The true left end of clone CTD-2530H13 is at 104498 in this
             sequence. The true right end of clone XX-BWXD718 is at 2000 in this
             sequence.
FEATURES             Location/Qualifiers
     source          1..106497
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q13.1-13.3"
                     /clone="RP11-449M9"
                     /clone_lib="RPCI-11.2"
     misc_feature    10905..11106
                     /note="Sequence from clone PCR only."
ORIGIN
```

```
Qy          931 ACAGGGGGTCTCAGGAGAGAGCCT 952
                |||||||||||||||||||||||
Db        19854 ACAGGGGGTCTCAGGAGAGAGCCT 19833
```

Query Match   2.1%; Score 22; DB 9; Length 106497;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
LOCUS   CQ869788   112453 bp   DNA   linear   PAT 13-SEP-2004
CQ869788/c
DEFINITION   Sequence 209 from Patent WO2004074320.
ACCESSION   CQ869788
VERSION   CQ869788.1 GI:51999587
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 112453)
AUTHORS   Morris,D.W., Morris,D.W. and Malandro,M.S.
TITLE   Novel therapeutic targets in cancer
JOURNAL   Patent: WO 2004074320-A 209 02-SEP-2004;
             Sagres Discovery, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..112453
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Qy          113 CTGTCTGAGGCCACACAGCTAG 134
                ||||||||||||||||||||||
```

Query Match   2.1%; Score 22; DB 6; Length 112453;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db       33503 CTGTCTGAGGCCACACAGCTAG 33482
             |||||||||||||||||||||
             |||||||||||||||||||||
```

RESULT 12
LOCUS   AC137629/c
DEFINITION AC137629
ACCESSION   Homo sapiens clone RP11-525N13, WORKING DRAFT SEQUENCE.
VERSION   AC137629
KEYWORDS   AC137629.16 GI:29501835
SOURCE   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
ORGANISM   Homo sapiens
            Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 125836)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooke,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,J., Newton,J., Nguyen,D., Nguyen,N., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Oknuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojukokun,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,H., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 125836)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 125836)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (05-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Apr 3, 2003 this sequence version replaced gi:29423847.

            Center: Baylor College of Medicine
            Genome Center
```

---

```
FEATURES             Location/Qualifiers
     source          1..125836
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-525N13"
     misc_feature    41073..71172
                     /note="remaining sized gap in near_complete"
     misc_feature    90658..90857
                     /note="remaining sized gap in near_complete"

ORIGIN
Db       56334 TGTCTGAGGCCACACAGCTAGA 56313
             ||||||||||||||||||||||
             ||||||||||||||||||||||
Qy         114 TGTCTGAGGCCACACAGCTAGA 135

RESULT 13
LOCUS       AC002419/c
ACCESSION   AC002419
DEFINITION  Homo sapiens chromosome X clone bWXD40, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC002419
VERSION     AC002419.2  GI:21405639
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128440)
AUTHORS     Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and
            Mazzarella,R.
TITLE       Direct Submission
JOURNAL     Unpublished (1997)
REFERENCE   2  (bases 1 to 128440)
AUTHORS     Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE       Direct Submission
JOURNAL     Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
```

```
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Drafting Center Code: BCM
             Contact: hgsc-help@bcm.tmc.edu
             ----------- Project Information
             Center project name: RP11-525N13
             Center clone name: RP11-525N13
             ----------- Summary Statistics
             Sequencing vector: Plasmid;
             Sequencing vector: M13;
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Sequencing quality: Phrap; version 0.990329
             Assembly program: Phrap; version 0.990329
             Consensus quality: 125202 bases at least Q40
             Consensus quality: 125396 bases at least Q30
             Consensus quality: 125495 bases at least Q20
             Estimated insert size: 136468; sum-of-contigs estimation
             Quality coverage: 20x in Q20 bases; sum-of-contigs estimation
             --------------------------------------------------------------
  *   NOTE: Estimated insert size may differ from sequence length
  *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
  *   NOTE: This is a 'near_complete' sequence.
  *   NOTE: This is a 'working draft' sequence. It currently
  *       consists of 1 contigs. Gaps between the contigs
  *       are represented as runs of N. The order of the pieces
  *       is believed to be correct as given, however the sizes
  *       of the gaps between them are based on estimates that have
  *       provided by the submitter.
  *   This sequence will be replaced
  *   by the finished sequence as soon as it is available and
  *   the accession number will be preserved.
  *   125836: contig of 125836 bp in length.
  *
```

```
Qy         114 TGTCTGAGGCCACACAGCTAG 135
               |||||||||||||||||||||
               |||||||||||||||||||||
Db       56334 TGTCTGAGGCCACACAGCTAG 56313
```

```
Query Match         2.1%;  Score 22;  DB 2;  Length 125836;
Best Local Similarity  100.0%;  Pred. No. 1.2;
Matches  22;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

COMMENT   Washington University School of Medicine, 4566 Scott Avenue, St.
           Louis, MO 63110, USA
           On Jun 13, 2002 this sequence version replaced gi:22323250.
           Current status of this project is available at:
           'http://genome.wustl.edu/cgm/seq_projects.html'
           Submitted by:
           Ellson Chen,
           Advanced Center for Genetic Technology,
           Applied Biosystems Division of Perkin Elmer Corp.,
           850 Lincoln Center Drive,
           Foster City, CA 94404 USA
           e-mail: ellson@geneq.apldbio.com

           and

           Buddy Brownstein,
           Center for Genetics in Medicine,
           Washington University School of Medicine, Box 8232
           4566 Scott Avenue,
           St. Louis, MO 63108 USA
           e-mail: buddy@genetics.wustl.edu

           and

           David J. States,
           Institute for Biomedical Computing
           Washington University in St. Louis
           700 South Euclid Ave.
           St. Louis, MO 63108 USA
           e-mail: states@ibc.wustl.edu.
           NOTE: This is a 'working draft' sequence. It currently
           consists of 2 contigs. The true order of the pieces
           is not known and their order in this sequence record is
           arbitrary. Gaps between the contigs are represented as
           runs of N, but the exact sizes of the gaps are unknown.
           This record will be updated with the finished sequence
           as soon as it is available and the accession number will
           be preserved.
           *
           *   1   58852: contig of 58852 bp in length
           *   58853   58852: gap of 100 bp
           *   58953   128440: contig of 69488 bp in length

FEATURES   Location/Qualifiers
     source   1..128440
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="DWXD40"

ORIGIN

Qy   931 ACAGGGGGTCTCAGGAAGAGCCT 952
              ||||||||||||||||||||||||
Db   103762 ACAGGGGGTCTCAGGAAGAGCCT 103741

RESULT 14
LOCUS   AL844513   131184 bp   DNA   linear   ROD 18-NOV-2002
DEFINITION Mouse DNA sequence from clone RP23-132A3 on chromosome 2, complete
            sequence.
ACCESSION   AL844513
VERSION   AL844513.8 GI:33438736
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 131184)
  AUTHORS   Johnson,C.

Query Match   2.1%; Score 22; DB 2; Length 128440;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   58852: contig of 58852 bp in length
          58852: gap of 100 bp
          128440: contig of 69488 bp in length

TITLE   Direct Submission
JOURNAL   Submitted (19-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Aug 5, 2003 this sequence version replaced gi:24816942.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence. Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk

            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-132A3 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.

FEATURES   Location/Qualifiers
     source   1..131184
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-132A3"
                     /clone_lib="RPCI-23"

ORIGIN

Qy   65 AGTCCCATTTTTCAGATGAGGA 86
             ||||||||||||||||||||||
Db   2737 AGTCCCATTTTTCAGATGAGGA 2758

RESULT 15
LOCUS   AC104081   142118 bp   DNA   linear   PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-559A1 from 2, complete sequence.
ACCESSION   AC104081
VERSION   AC104081.2 GI:17977480
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 142118)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792

Query Match   2.1%; Score 22; DB 10; Length 131184;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
PUBMED      9847074
REFERENCE   2  (bases 1 to 142118)
AUTHORS     Isak,A. and Meyer,R.
TITLE       The sequence of Homo sapiens BAC clone RP11-559A1
JOURNAL     Unpublished (2001)
REFERENCE   3  (bases 1 to 142118)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (03-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 142118)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (22-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 142118)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 142118)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Dec 22, 2001 this sequence version replaced gi:17227302.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            --------------  Summary Statistics
            Center project name: H_NH0559A01
            Drafting Center: WIBR
            --------------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-50B16; the clone sequenced
            to the right is RP11-554H10, 2000 bp overlap. Actual end of this
            clone is at base position 79619 of RP11-554H10.
```

```
FEATURES             Location/Qualifiers
     source          1..142118
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-559A1"
                     /clone_lib="RPCI-11"

     repeat_region   1..964
                     /rpt_family="L1"
     repeat_region   971..1204
                     /rpt_family="L1"
     repeat_region   976..1036
                     /rpt_family="L1"
     repeat_region   1073..1114
                     /rpt_family="AT_rich"
     repeat_region   1073..1114
                     /rpt_family="(TATAA)n"
     repeat_region   1243..1518
                     /rpt_family="MER1_type"
     repeat_region   1271..1338
                     /rpt_family="AT_rich"
     repeat_region   1553..1580
                     /rpt_family="AT_rich"
     repeat_region   1814..1891
                     /rpt_family="L2"
     repeat_region   2254..2741
                     /rpt_family="MaLR"
     repeat_region   3256..3311
                     /rpt_family="MIR"
     repeat_region   3370..6404
                     /rpt_family="L2"
     repeat_region   4153..4187
                     /rpt_family="AT_rich"
     repeat_region   4234..4317
                     /rpt_family="ERVL"
     repeat_region   4750..4793
                     /rpt_family="L2"
     repeat_region   5564..5620
                     /rpt_family="AT_rich"
     repeat_region   6370..6404
                     /rpt_family="MIR"
     repeat_region   6550..6681
                     /rpt_family="(CA)n"
     repeat_region   6704..7158
                     /rpt_family="MIR"
     repeat_region   7009..7031
                     /rpt_family="ERVL"
     repeat_region   7189..7354
                     /rpt_family="(T)n"
     repeat_region   7650..7692
                     /rpt_family="(ATG)n"
     repeat_region   7929..7996
                     /rpt_family="AT_rich"
     repeat_region   7984..8103
                     /rpt_family="MIR"
     repeat_region   8425..8461
                     /rpt_family="MIR"
     repeat_region   8543..8588
                     /rpt_family="MIR"
     repeat_region   8637..9095
                     /rpt_family="MIR"
     repeat_region   9178..9260
                     /rpt_family="MaLR"
     repeat_region   9684..9808
                     /rpt_family="MIR"
     repeat_region   10301..10341
                     /rpt_family="MIR"
     repeat_region   10978..11002
                     /rpt_family="(CA)n"
     repeat_region   11305..11402
                     /rpt_family="AT_rich"
     repeat_region   11305..11402
                     /rpt_family="MIR"
```

Data from AC012671 was used to finish this clone, AC0668259.
Polymorphisms have been identified between AC012671 and AC068259.
The sequence of AC068259 has been incorporated into AC104081.

```
repeat_region    12182..12244
                 /rpt_family="Alu"
repeat_region    12245..12587
                 /rpt_family="MaLR"
repeat_region    14154..14485
                 /rpt_family="MaLR"
repeat_region    14486..14725
                 /rpt_family="Alu"
repeat_region    14704..14749
                 /rpt_family="Alu"
repeat_region    15138..15429
                 /rpt_family="(TAAA)n"
repeat_region    15811..16068
                 /rpt_family="Alu"
repeat_region    16069..16458
                 /rpt_family="MaLR"
repeat_region    16459..16530
                 /rpt_family="MER1_type"
repeat_region    17696..18200
                 /rpt_family="MER1_type"
repeat_region    18458..18597
                 /rpt_family="ERV1"
repeat_region    18598..19184
                 /rpt_family="MaLR"
repeat_region    18943..18997
                 /rpt_family="MaLR"
repeat_region    19185..19515
                 /rpt_family="GA-rich"
repeat_region    19593..20149
                 /rpt_family="MaLR"
repeat_region    20252..20416
                 /rpt_family="L2"
repeat_region    20842..20881
                 /rpt_family="Alu"
repeat_region    20853..21166
                 /rpt_family="(TTTTC)n"
repeat_region    21410..21551
                 /rpt_family="MIR"
repeat_region    22386..22535
                 /rpt_family="CT-rich"
```

```
Query Match       2.1%;   Score 22;   DB 9;  Length 142118;
Best Local Similarity  100.0%;  Pred. No. 1.2;
Matches   22;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy        65 AGTCCCATTTTTCAGATGAGGA 86
             ||||||||||||||||||||||
Db     24601 AGTCCCATTTTTCAGATGAGGA 24622
```

Search completed: June 29, 2005, 19:59:07
Job time : 4895 secs

THIS PAGE BLANK (USPTO)